

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 12:15:47 ; Search time 103.919 Seconds
(without alignments)
1035.886 Million cell updates/sec

Title: US-10-073-293A-4
Perfect score: 1248
Sequence: 1 MESTTPQAPGSAFMFGCK.....VCGCLTALIOAFWQGPDEL 245

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*
1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001s.*
5: Geneseq2002s.*
6: Geneseq2003as.*
7: Geneseq2003bs.*
8: Geneseq2004s.*
9: Geneseq2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1248	100.0	245	5	ABP57756 E. coli L
2	1248	100.0	245	9	ADZ39951 E.coli b2
3	1248	100.0	245	9	ADZ36195 Escherich
4	1013.5	81.2	251	7	ABO63670 Klebsiell
5	841.5	67.4	261	6	ABM69066 Photornab
6	592	47.4	266	6	ADA35238 Acinetoba
7	246.5	19.8	247	7	ABO61961 Klebsiell
8	229.5	18.4	230	5	AAG81587 S. epider
9	228	18.3	235	5	ABB47959 Listeria
10	225.5	18.1	242	8	ABP40309 Staphyloc
11	225.5	18.1	242	8	ADS07426 Staphyloc
12	219.5	17.6	230	5	ABP30530 Streptoco
13	219.5	17.6	230	8	ADV81752 Streptoco
14	219.5	17.6	231	8	ABP28562 Streptoco
15	219.5	17.6	231	8	ADV88332 Streptoco
16	219.5	17.6	231	8	ADV79585 Streptoco
17	214.5	17.2	255	7	ADC95293 E. faeciu
18	209	16.4	345	7	ABO79201 Pseudomon
19	204.5	16.4	238	6	ABM72815 Staphyloc
20	204	16.3	235	5	ABBS4959 Lactococc
21	199.5	16.0	263	6	ADA33199 Acinetoba
22	197	15.8	251	4	ABM86247 C. glutam
23	197	15.8	251	4	AAG90035 C glutami
24	194.5	15.6	206	7	ADH88904 Enterococ

25 189 15.1 226 6 ADB08660
26 187 15.0 224 8 ADK48319
27 180 14.4 237 4 AAG93154
28 178 14.3 218 6 ABU00481
29 178 14.3 218 6 ABP81300
30 178 14.3 218 8 ADM92060
31 172 13.8 214 6 ADB08658
32 171 13.7 185 8 ADR94706
33 171 13.7 185 9 AEA58576
34 167.5 13.4 338 5 ABP66261
35 146.5 11.7 230 7 ADF03888
36 143 11.5 288 7 ABO65701
37 122.5 9.8 247 6 ADA34737
38 113.5 9.1 112 3 AAY74752
39 113.5 9.1 202 9 ABM94911
40 107.5 8.6 397 6 ABM72496
41 106 8.5 113 3 AAY74753
42 105 8.4 309 6 ABM68599
43 105 8.4 324 4 AAG91133
44 105 8.4 324 7 ADL65433
45 103.5 8.3 432 1 AAP70459

ALIGNMENTS

RESULT 1
ABP57756
ID ABP57756 standard; protein; 245 AA.
XX
AC ABP57756;
XX
DT 29-JAN-2003 (first entry)
XX
DE E. coli L-amino acid producing protein #1.
XX
KW L-amino acid; E. coli.
XX
OS Escherichia coli.
XX
PN EP1239041-A2.
XX
PD 11-SEP-2002.
XX
PF 13-FEB-2002; 2002EP-00003335.
XX
PR 13-FEB-2001; 2001RU-00103865.
PR 26-FEB-2001; 2001RU-00104998.
PR 26-FEB-2001; 2001RU-00104999.
PR 28-JUN-2001; 2001RU-00117632.
PR 28-JUN-2001; 2001RU-00117633.
XX (AJIN) AJINOMOTO CO INC.
XX
XX Tabolina EA, Rybak KV, Khourges EM, Voroshilova EB, Gussyatiner MM;
XX WPI; 2002-699856/76.
XX N-PSDB; ABV75656.

Novel L-amino acid producing Escherichia bacterium, is modified to enhance L-amino acid production by enhancing the activities of protein capable of making bacterium to have enhanced resistance to L-amino acids.
Disclosure; Page 23; 33pp; English.
The invention relates to a novel L-amino acid producing bacterium of genus Escherichia, modified to enhance L-amino acid production by enhancing the activities of a protein. The novel bacterium is useful for producing L-amino acid e.g. L-Thr, L-Pro, L-Met and L-Arg, by cultivating the bacterium in a culture medium and collecting L-amino acid to be produced and accumulated from the culture medium. The present sequence represents a protein of the invention which causes increased L-amino acid production in E. coli

XX Sequence 245 AA;
 SQ Query Match 100.0%; Score 1248; DB 5; Length 245;
 Best Local Similarity 100.0%; Pred. No. 8.8e-122;
 Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MESPTQAPGSAATFMEGCKDSLPIVISYIPVAFGLNATRLGFSPLSVFFSCIIYAG 60
 Db 1 MESPTQAPGSAATFMEGCKDSLPIVISYIPVAFGLNATRLGFSPLSVFFSCIIYAG 60

Qy 61 ASQFVITAMLAAGSSLTAAITVMAMDVHRVLYGPSLSRSRIIQRLQSKTALMAFGLTDE 120
 Db 61 ASQFVITAMLAAGSSLTAAITVMAMDVHRVLYGPSLSRSRIIQRLQSKTALMAFGLTDE 120

Qy 121 VFAAATAKLVNRNRWSENMMIGIAFSSWSWVFGTVIGAFSGSLGQYPAVEAALGFM 180
 Db 121 VFAAATAKLVNRNRWSENMMIGIAFSSWSWVFGTVIGAFSGSLGQYPAVEAALGFM 180

Qy 191 LPALFMSFLLASFORQSLCVTAALVGALAGVTLFSIPVAILAGIVCGCLTALIOAFWQG 240
 Db 181 LPALFMSFLLASFORQSLCVTAALVGALAGVTLFSIPVAILAGIVCGCLTALIOAFWQG 240

Qy 241 APDEL 245
 Db 241 APDEL 245

RESULT 2
 ADZ39951
 ID ADZ39951 standard; protein; 245 AA.
 XX AC ADZ39951;
 XX 14-JUL-2005 (first entry)
 DE E.coli b2682 protein for improved amino acid production in bacteria.
 XX amino acid production; fermentation; transport protein.
 XX Escherichia coli.
 XX EP1526179-A1.
 XX 27-APR-2005.
 XX 13-FEB-2002; 2004EP-00028876.
 XX 13-FEB-2001; 2001RU-00103865.
 XX 26-FEB-2001; 2001RU-00104998.
 XX 28-JUN-2001; 2001RU-00117632.
 XX 28-JUN-2001; 2001RU-00117632.
 XX 13-FEB-2002; 2002EP-00003335.
 XX (AJIN) AJINOMOTO CO INC.
 XX Tabolina EA, Rybak KV, Khourges EM, Voroshilova EB, Gusyatiner MM;
 XX WPI; 2005-308098/32.
 XX N-PSDB; ADZ39950.
 XX New L-amino acid producing bacterium belonging to the genus Escherichia,
 XX useful for producing L-amino acids, e.g. L-threonine, L-valine, L-
 XX proline, L-methionine, or L-arginine.
 XX Disclosure; SEQ ID NO 4; 35pp; English.
 XX The invention relates to an L-amino acid producing bacterium belonging to
 XX the genus Escherichia, where the bacterium has been modified so that the
 XX L-amino acid production by the bacterium is enhanced by enhancing
 XX activities of proteins by transformation of the bacterium with DNA coding
 XX for protein or by alteration of an expression regulation sequence of the

CC DNA on the chromosome of the bacterium. The L-amino acid producing
 CC bacterium is useful for producing L-amino acids by fermentation,
 CC including L-threonine, L-valine, L-proline, L-methionine, or L-arginine.
 CC The genes are useful for improving L-amino acid productivity. This
 CC sequence corresponds to the b2682 protein, one of the L-amino acid
 CC biosynthesis pathway proteins. The protein is a putative transport
 CC protein within the biosynthesis pathway.

XX Sequence 245 AA;
 SQ Query Match 100.0%; Score 1248; DB 9; Length 245;
 Best Local Similarity 100.0%; Pred. No. 8.8e-122;
 Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MESPTQAPGSAATFMEGCKDSLPIVISYIPVAFGLNATRLGFSPLSVFFSCIIYAG 60
 Db 1 MESPTQAPGSAATFMEGCKDSLPIVISYIPVAFGLNATRLGFSPLSVFFSCIIYAG 60

Qy 61 ASQFVITAMLAAGSSLTAAITVMAMDVHRVLYGPSLSRSRIIQRLQSKTALMAFGLTDE 120
 Db 61 ASQFVITAMLAAGSSLTAAITVMAMDVHRVLYGPSLSRSRIIQRLQSKTALMAFGLTDE 120

Qy 121 VFAAATAKLVNRNRWSENMMIGIAFSSWSWVFGTVIGAFSGSLGQYPAVEAALGFM 180
 Db 121 VFAAATAKLVNRNRWSENMMIGIAFSSWSWVFGTVIGAFSGSLGQYPAVEAALGFM 180

Qy 181 LPALFMSFLLASFORQSLCVTAALVGALAGVTLFSIPVAILAGIVCGCLTALIOAFWQG 240
 Db 181 LPALFMSFLLASFORQSLCVTAALVGALAGVTLFSIPVAILAGIVCGCLTALIOAFWQG 240

Qy 241 APDEL 245
 Db 241 APDEL 245

RESULT 3
 ADZ36195
 ID ADZ36195 standard; protein; 245 AA.
 XX AC ADZ36195;
 XX 14-JUL-2005 (first entry)
 DE Escherichia coli b2682 protein SEQ ID NO:4.
 XX amino acid production.
 XX Escherichia coli.
 XX EP1526181-A1.
 XX 27-APR-2005.
 XX 13-FEB-2002; 2004EP-00028877.
 XX 13-FEB-2001; 2001RU-00103865.
 XX 26-FEB-2001; 2001RU-00104998.
 XX 28-JUN-2001; 2001RU-00117632.
 XX 28-JUN-2001; 2001RU-00117632.
 XX 13-FEB-2002; 2002EP-00003335.
 XX (AJIN) AJINOMOTO CO INC.
 XX Tabolina EA, Rybak KV, Khourges EM, Voroshilova EB, Gusyatiner MM;
 XX WPI; 2005-317157/33.
 XX N-PSDB; ADZ36194.
 XX New modified L-amino acid producing bacterium useful to enhance the
 XX production of L-amino acid by enhancing the activities of proteins in a
 XX cell of bacterium.

PS Disclosure; SEQ ID NO 4; 35pp; English.
 XX
 CC The invention relates to an L-amino acid (A) producing bacterium (I) (belonging to the genus Escherichia), which is modified to enhance the production of (A) by enhancing the activities of proteins (G) or (H) in a cell of (I). Also described is a method for producing (A) comprising culturing the bacterium in a culture medium and collecting the produced and accumulated L-amino acid. The modified bacterium has the ability to grow on a minimal medium containing L-amino acid or its analog in a minimal concentration and ability to grow faster on a medium containing L-amino acid or its analog than the unmodified strain or the wild type strain, or the parental strain of the bacterium. The present sequence represents the E. coli b2862 protein which has L-amino acid excretion activity.

Query Match 100.0%; Score 1248; DB 9; Length 245;
 Best Local Similarity 100.0%; Pred. No. 8.8e-122;
 Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 SQ Sequence 245 AA;
 QY 1 MESPTPOAGSATFMEGCKDSLPIVISYIPVAFGLNATRLGFSPLSVFFSCIIYAG 60
 DB 1 MESPTPOAGSATFMEGCKDSLPIVISYIPVAFGLNATRLGFSPLSVFFSCIIYAG 60
 QY 61 ASQFVITAMLAAGSSLMIAALTVMAMDVHRVLYGFSLSRSRIIQRLOKSKTALWAFGLTDE 120
 DB 61 ASQFVITAMLAAGSSLMIAALTVMAMDVHRVLYGFSLSRSRIIQRLOKSKTALWAFGLTDE 120
 QY 121 VFPAATAKLVRNRRWSENMMIGIAFSSWSWVFGTVIGAPSGGLLQGYPAVEAALGFM 180
 DB 121 VFPAATAKLVRNRRWSENMMIGIAFSSWSWVFGTVIGAPSGGLLQGYPAVEAALGFM 180
 QY 181 LPALFMSFLLASFORKQSLCVTAALVGLAGVTLFSPVAILAGIVCGCLTALIQAFWQ 240
 DB 181 LPALFMSFLLASFORKQSLCVTAALVGLAGVTLFSPVAILAGIVCGCLTALIQAFWQ 240
 QY 241 APDEL 245
 DB 241 APDEL 245

RESULT 4
 ABO63670
 ID ABO63670 standard; protein; 251 AA.
 AC ABO63670;
 DT 29-JUL-2004 (first entry)
 DE Klebsiella pneumoniae polypeptide seqid 10187.
 KW Recombinant expression vector; transcription regulatory element;
 KW Klebsiella pneumoniae protein; antibacterial; vaccine.
 OS Klebsiella pneumoniae.
 PN US6610836-B1.
 XX 26-AUG-2003.
 XX 27-JAN-2000; 2000US-00489039.
 XX 29-JAN-1999; 99US-0117747P.
 XX (GENO-) GENOME THERAPEUTICS CORP.
 PA Breton GL, Osborne M;
 PI WPI; 2003-895346/82.
 DR N-PSDB; ACH97221.
 XX New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for

PT preparing a vaccine composition against Klebsiella pneumoniae.
 PS Disclosure; SEQ ID NO 10187; 932pp; English.
 XX
 CC The invention describes a new isolated nucleic acid encoding a Klebsiella pneumoniae polypeptide. Also described are: a recombinant expression vector comprising the nucleic acid, operably linked to a transcription regulatory element; and a cell comprising the recombinant expression vector. The nucleic acid is useful for preparing a vaccine composition against Klebsiella pneumoniae. This is the amino acid sequence of a Klebsiella pneumoniae polypeptide of the invention
 SQ Sequence 251 AA;

Query Match 81.2%; Score 1013.5; DB 7; Length 251;
 Best Local Similarity 81.0%; Pred. No. 3e-97;
 Matches 200; Conservative 18; Mismatches 26; Indels 3; Gaps 1;
 QY 1 MESPTPOQA---PGSATFMEGCKDSLPIVISYIPVAFGLNATRLGFSPLSVFFSCII 57
 DB 4 MENPAPLTCALPERVATVGEVGDLSLPIVISYIPVAFGLNATRLGFTPLSLFFSCII 63
 QY 58 YAGASQVITAMLAAGSSLMIAALTVMAMDVHRVLYGFSLSRSRIIQRLOKSKTALWAFGL 117
 DB 64 YAGASQVITAMLAAGSSLMIAALTVMAMDVHRVLYGFSLSRSRIIQRLOKSKTALWAFGL 123
 QY 118 TDEVFAATAKLVRNRRWSENMMIGIAFSSWSWVFGTVIGAPSGGLLQGYPAVEAAL 177
 DB 124 TDEVFAATAKLVRNRRWSENMMIGIAFSSWSWVFGTVIGAPSGGLLQGYPAVEAAL 183
 QY 178 GFMLPALFMSFLLASFORKQSLCVTAALVGLAGVTLFSPVAILAGIVCGCLTALIQAF 237
 DB 184 GFMLPALFMSFLLASFORKQSLCVTAALVGLAGVTLFSPVAILAGIVCGCLTALIQAF 243
 QY 238 WQGAPE 244
 DB 244 LKGMPE 250

RESULT 5
 ABM69066
 ID ABM69066 standard; protein; 261 AA.
 AC ABM69066;
 DT 20-NOV-2003 (first entry)
 DE Photorhabdus luminescens protein sequence #2163.
 KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
 KW detection; food; gene expression; plant; animal; microorganism; toxin;
 KW antibiotic; biopesticide; virulence factor; disease model; plague;
 KW whooping cough.
 OS Photorhabdus luminescens.
 PN WO200294867-A2.
 XX 28-NOV-2002.
 XX 07-FEB-2002; 2002WO-IB003040.
 XX 07-FEB-2001; 2001FR-00001659.
 XX (INSP) INST PASTEUR.
 PA (CNRS) CNRS CENT NAT RECH SCI.
 XX
 XX Duchaud B, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
 XX Buchrieser C;
 DR WPI; 2003-148459/14.
 XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,

useful e.g. as therapeutic antimicrobials and agricultural pesticides.

Claim 2; SEQ ID NO 2163; 1205pp; French.

The invention relates to the isolation of genes and their encoded proteins from *Photobacterium luminescens*. The isolated sequences are sources of probes and primers for detecting the genome of *P. luminescens* and related species; to study polymorphisms, for gene analysis and for detection/amplification of the genes. Antibodies (Ab) raised against the polypeptides encoded by the genes are used for detection/identification of *P. luminescens*, e.g. in foods. The genes, proteins, Ab and cells that carry a gene-containing vector are used to select compounds that modulate, regulate, induce or inhibit expression of the genes in plants, animals or microorganisms other than *P. luminescens* and are able to alter response or sensitivity to toxins and antibiotics produced by *P. luminescens*. Cells transformed to express the genes are useful for recombinant production of the proteins, particularly toxins and antibacterials useful as insecticides, bactericides and fungicides. The genes, proteins, vectors containing the genes and Ab are also useful therapeutically (to treat microbial infection by bacteria or fungi that are sensitive to *P. luminescens*-encoded toxins or antibiotics) and as biopesticides. Other uses of the genes and the proteins are as virulence factors and for identifying targets of human diseases for which *P. luminescens* is a model (particularly plague and whooping cough). This sequence represents one of the isolated *P. luminescens* proteins

Sequence 261 AA;

Query Match 67.4%; Score 841.5; DB 6; Length 261;
Best Local Similarity 68.4%; Pred. No. 3e-79;
Matches 160; Conservative 32; Mismatches 41; Indels 1; Gaps 1;
QY 3 SPTQPARG-SATFMGCKDSLPIVISYIPVAFGLNATRLGFSPLSVFSCIIYAGA 61
DB 12 SDTSSPLTSKKSSETEGIIDSLPIVIGYIPVAFGLNATRLGFSPLSVFSCIIYAGA 71
QY 62 SQFVTAMLAAGSSILWIAALTVMAMDVRHLYGPSLRSRIIRLQKSKTALWAFGLTDEV 121
DB 72 SQFVTALLSAGTSLWISALTVMAMDVRHLYGPSLRSRIIRLQKSKTALWAFGLTDEV 131
QY 122 FAATAKLVRNRRNSENWMIIGIAFSSWSSVFGTVIGAFSGGLLQGYPAVEAALGFML 181
DB 132 FAATAKLVRNRRNSENWMIIGIAFSSWSSVFGTVIGAFSGGLLQGYPAVEAALGFML 191
QY 182 PALFMSFLASCRKQNSCVATLTGALTGITFFSIPVAILAGIVCGCLTALIQ 235
DB 192 PALFMSFLASCRKQNSCVATLTGALTGITFFSIPVAILAGIVCGCLTALIQ 245

RESULT 6
ADA35238
ID ADA35238 standard; protein; 266 AA.
AC ADA35238;
XX
XX
XX

20-NOV-2003 (first entry)

Acinetobacter baumannii protein #2399.

Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;
plant biocontrol agent.

Acinetobacter baumannii.

US6562958-B1.

13-MAY-2003.

04-JUN-1999; 99US-00328352.

09-JUN-1998; 98US-0088701P.

(GENO-) GENOME THERAPEUTICS CORP.

Breton G, Bush D;
WPI; 2003-576092/54.
N-PSDB; ADA31112.

New Acinetobacter baumannii proteins and nucleic acids, useful as reagents for diagnosing a bacterial disease, as components of antibacterial vaccines, as targets for antibacterial drugs, or as biocontrol agents for plants.

Example; SEQ ID NO 6525; 328pp; English.

The invention relates to isolated Acinetobacter baumannii nucleic acids. The A. baumannii nucleic acids and polypeptides are useful as reagents for diagnosing a bacterial disease, as components of antibacterial vaccines, as targets for antibacterial drugs, to detect the presence of A. baumannii and other Acinetobacter species in a sample, in screening compounds for the ability to interfere with the A. baumannii life cycle or to inhibit A. baumannii infection, and as biocontrol agents for plants. The present sequence represents the amino acid sequence of an A. baumannii protein.

Sequence 266 AA;

Query Match 47.4%; Score 592; DB 6; Length 266;
Best Local Similarity 48.1%; Pred. No. 3.8e-53;
Matches 111; Conservative 53; Mismatches 67; Indels 0; Gaps 0;
QY 7 QPAPGSATFMGCKDSLPIVISYIPVAFGLNATRLGFSPLSVFSCIIYAGASQFVI 66
DB 26 QMAAQATATFWGAKDSQAIVLTLPVSFAFGVSASQFGFTPWEAFLLSCSMYAGASQPLV 85
QY 67 TAMLAAGSSILWIAALTVMAMDVRHLYGPSLRSRIIRLQKSKTALWAFGLTDEVFAAT 126
DB 86 VALLASGSSILWLTALTVTALDIRHLYGPALYNLPTKLNKKTAVWAWGLTDEVFASGM 145
QY 127 AKLVNRRNRRNSENWMIIGIAFSSWSSVFGTVIGAFSGGLLQGYPAVEAALGFMLPALFM 186
DB 146 IQLSQRQKQWSESWMGLGSLFSPNWSWASGSLGLGFADQVAHLKFLQALDFLLPALFL 205
QY 187 SFLIASFQRKQSLCVTAALVAGALAGVTLSIPVAILAGIVCGCLTALIQAF 237
DB 206 SFLIAAFERKSLVAVSLVVSALACYWINLSAAIFIGILSGILAGLPKY 256

RESULT 7
ABO61961
ID ABO61961 standard; protein; 247 AA.
XX ABO61961;
AC ABO61961;
XX
XX
XX

29-JUL-2004 (first entry)

Klebsiella pneumoniae polypeptide seqid 8478.

Recombinant expression vector; transcription regulatory element;
Klebsiella pneumoniae protein; antibacterial; vaccine.

Klebsiella pneumoniae.

US6610836-B1.

26-AUG-2003.

27-JAN-2000; 2000US-00489039.

29-JAN-1999; 99US-0117747P.

(GENO-) GENOME THERAPEUTICS CORP.

Breton GL, Osborne M;

XX

DR	WPI, 2003-895346/82.
DR	N-PSDB; ACH95512.
XX	
PT	New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for
PT	preparing a vaccine composition against Klebsiella pneumoniae.
XX	
XX	Disclosure; SEQ ID NO 8478; 932pp; English.
XX	
CC	The invention describes a new isolated nucleic acid encoding a Klebsiella
CC	pneumoniae polypeptide. Also described are: a recombinant expression
CC	vector comprising the nucleic acid, operably linked to a transcription
CC	regulatory element; and a cell comprising the recombinant expression
CC	vector. The nucleic acid is useful for preparing a vaccine composition
CC	against Klebsiella pneumoniae. This is the amino acid sequence of a
CC	Klebsiella pneumoniae polypeptide of the invention
XX	
XX	Sequence 247 AA;
XX	
XX	Query Match 19.8%; Score 246.5; DB 7; Length 247;
XX	Best Local Similarity 28.6%; Pred. No. 4.5e-17;
XX	Matches 67; Conservative 46; Mismatches 106; Indels 15; Gaps 6;
XX	
QY	18 GKCDLPTIVISYIPVAPAFGLNATRLGSPLESVFSCIIYAGASQFVITAWLAAGSSLW 77
Db	10 GASAINPLCIGDFFPSFIVGALSVSAGSMVQSTAWSAIVTAGSAQMLANMLTKGTATLG 69
XX	
QY	78 IAAITVMAMDVHRVLYGSPLSRSRIIQRLOKSKTALWAFGLTDEVFAAATAKLVRNRRWS 137
Db	70 VIIFTLLINLRHLYLSASI-SCTVREASFFKCFMSYALTDEVY-ATTVKEMEGNKKEK 127
XX	
QY	138 ENNMIGTAFSSWSNWFTGTVIGFSGGLGQYPAVEA-ALGFMLPALFMSFLLASFORK 196
Db	128 YLFYGSAMITFWAIWLADFLGALVGA----SFPHEIKYGLDFAMVAAFIAIVVP--QIK 181
XX	
QY	197 QSLCVTAALVCALAGVTLFSP-----VAILAGIVCGCLTALIQAFWOGAPDE 244
Db	182 SQACTVAAVAASVGVLLWLPYSLGIVASVLGVLGACVLDLAERKQMAKTE 235
XX	
XX	RESULT 8
XX	AAG81587
ID	AAG81587 standard; protein; 230 AA.
XX	AAG81587;
XX	
DT	03-SEP-2001 (first entry)
XX	
DE	S. epidermidis open reading frame protein sequence SEQ ID NO:268.
XX	
KW	Staphylococcus epidermidis SRI strain; infection; diagnosis; vaccination;
KW	endocarditis.
XX	
OS	Staphylococcus epidermidis.
XX	
PN	WO200134809-A2.
XX	
PD	17-MAY-2001.
XX	
XX	09-NOV-2000; 2000WO-US030782.
PF	
XX	
XX	09-NOV-1999; 99US-0164258P.
PR	
XX	(GLAX) GLAXO GROUP LTD.
PA	
PA	Kimmerly WJ;
PI	
XX	WPI; 2001-316495/33.
DR	N-PSDB; AAH52437.
XX	
PT	Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT	useful for vaccinating against infections, e.g. endocarditis.
XX	
XX	Claim 18; Page 114; 2188pp; English.
PS	

PN US2004147734-A1.
 XX 29-JUL-2004.
 XX 01-DEC-2003; 2003US-00724972.
 XX 08-NOV-1997; 97US-0064964P.
 XX 13-AUG-1998; 98US-00134001.
 XX 29-NOV-1999; 99US-00450969.
 XX (DOUC/) DOUCETTE-STAMM L.
 XX (BUSH/) BUSH D.
 XX Doucette-Stamm L, Bush D;
 XX WPI; 2004-580138/56.
 XX N-PSDB; ADS03654.
 XX New isolated polypeptide and encoding nucleic acid derived from
 PT Staphylococcus epidermidis, useful for diagnosing, preventing and/or
 PT treating an S. epidermidis bacterial infection.
 XX Claim 17; SEQ ID NO 6721; 741pp; English.
 XX The invention describes an isolated nucleic acid comprising a nucleotide
 CC sequence with any of 3772 fully defined nucleotide sequences (SEQ ID NO:
 CC 1-3772) and encoding an Staphylococcus epidermidis polypeptide with any
 CC of 3772 fully defined amino acid sequences (SEQ ID NO: 3772-7544) as
 CC given in the specification. Also described are: a recombinant expression
 CC vector; a cell comprising a recombinant expression vector of (1);
 CC producing an S. epidermidis polypeptide; an isolated nucleic acid
 CC comprising a nucleotide sequence of at least 8 nucleotides in length; a
 CC vaccine composition for prevention or treatment of an S. epidermidis
 CC infection, comprising a nucleic acid cited above and a carrier; treating
 CC a subject for S. epidermidis infection; a recombinant or substantially
 CC pure preparation of an S. epidermidis polypeptide or its fragment; a
 CC vaccine composition for prevention or treatment of an S. epidermidis
 CC infection; detecting the presence of a Staphylococcus nucleic acid in a
 CC sample; a computer readable medium having recorded in it the nucleotide
 CC sequences with SEQ ID NO: 1-3772 or its fragments; a computer based
 CC system for identifying fragments of the Staphylococcus genome of
 CC commercial importance; a computer based system for identifying fragments
 CC of the Staphylococcus plasmids of commercial importance; identifying
 CC commercially important nucleic acid fragments of the Staphylococcus
 CC genome and/or plasmids; and identifying an expression modulating fragment
 CC of the Staphylococcus genome and/or plasmids. The methods and
 CC compositions of the present invention are useful for the diagnosis,
 CC prevention and/or treatment of an Staphylococcus epidermidis bacterial
 CC infection. This is the amino acid sequence of a S. epidermis protein of
 CC the invention.
 XX Sequence 242 AA;
 XX
 Query Match 18.1%; Score 225.5; DB 8; Length 242;
 Best Local Similarity 28.6%; Pred. No. 6.9e-15;
 Matches 66; Conservative 39; Mismatches 89; Indels 37; Gaps 7;
 QY 14 TMEGCKDSLPIVTSYIPVAFAGLNATRLGFSPLSFVFFSCIIYAGASQFVITAMLAAG 73
 Db 17 TFKQGVKECTPLIGYAGVGLSPGIVAAASQNFVLEIILICLIYAGAAQFIICTLVIAG 76
 QY 74 SSLWIAALTVMADMVRHVLVGPISLRRIIQRLOKSKTALWAFGLTDEVFMAATAKLVNN 133
 Db 77 TPISAIVLTLLIVNSRFLLSMTLAPNKKYQGFNRVGLCTL-LTDETFGVAITPYVK-G 134
 QY 134 RRNSENMIGIAFSSWSWVFGTVIGAFSGSLGQYPAVEAIG--FMLPALFMSFLA 191
 Db 135 EKINDRWLHGLNITAYLFTWVSCVIGAIFGE-----YISNPDALGLDFAITAMFIFLCIS 189
 QY 192 SFQ--RKQSL-----CV-----TAALYCALAGVTL 214
 Db 190 QFEGIKRSRRIYIIVLCVIVMMLLSLILPSVAILIAIYAALIGVVM 240

RESULT 12
 ABP30530
 ID ABP30530 standard; protein; 230 AA.
 XX AC ABP30530;
 XX 02-JUL-2002 (first entry)
 XX Streptococcus polypeptide SEQ ID NO 10236.
 XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 KW group A streptococcus; Streptococcus pyogenes; antibacterial;
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
 XX OS Streptococcus agalactiae.
 XX WO200234771-A2.
 XX 02-MAY-2002.
 XX 29-OCT-2001; 2001WO-GB004789.
 XX 27-OCT-2000; 2000GB-00026333.
 XX 24-NOV-2000; 2000GB-00028727.
 XX 07-MAR-2001; 2001GB-00005640.
 XX (CHIR-) CHIRON SPA
 XX (GENO-) INST GENOMIC RES.
 XX Telford J, Maignani V, Margarit Y Rosi, Grandi G, Fraser C;
 PI Tettelin H;
 XX WPI; 2002-352536/38.
 XX N-PSDB; ABN71161.
 XX New Streptococcus protein for the treatment or prevention of infection or
 PT disease caused by Streptococcus bacteria, such as meningitis, and for
 PT detecting a compound that binds to the protein.
 XX Claim 1; Page 4153; 4525pp; English.
 XX The invention relates to a protein (ABP25413-ABP30895) from group B
 CC streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
 CC the specification. The proteins have antibacterial and antiinflammatory
 CC activity. (1), nucleic acids encoding (1), ABN6044-ABN71526 and
 CC antibodies that bind (1) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (1) are used to detect Streptococcus in a
 CC biological sample. (1) is used to determine whether a compound binds to
 CC (1). A composition comprising (1) or a nucleic acid encoding (1), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (1) may be used to recombinantly produce (1) and may be
 CC used in gene therapy. Antibodies to (1) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins
 XX Sequence 230 AA;
 XX
 Query Match 17.6%; Score 219.5; DB 5; Length 230;
 Best Local Similarity 29.0%; Pred. No. 2.8e-14;
 Matches 67; Conservative 38; Mismatches 87; Indels 39; Gaps 7;
 QY 15 FMEGCKDSLPIVTSYIPVAFAGLNATRLGFSPLSFVFFSCIIYAGASQFVITAMLAAGS 74
 Db 5 FKEGVKDALPTALGYISIGLAFGIVASASDLSAIEVGLMSALVYGSAQFAMCALLAKA 64
 QY 75 SLWIAALTVMADMVRHVLVGPISLRRIIQRLOKSKTALWAFG--LTDEVFMAATAKLVNN 132
 Db 65 DLMTITMTVFLNLRNLM--SLHATTIFKSAHNMNLQ-LAIGTLITDESYGVLLEAL-H 120

[illegible]

ADV88332;
AC
XX

This Page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 12:18:47 ; Search time 33.0337 Seconds
(without alignments)
713.608 Million cell updates/sec

Title: US-10-073-293A-4
Perfect score: 1248
Sequence: 1 MESPTPOPAPGSATFMEGCK.....VCGCLTALIOAFWQGAPDEL 245
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80.*
1: Pirl.*
2: Pirl2.*
3: Pirl3.*
4: Pirl4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1248	100.0	245	1 C65048	hypothetical prote
2	1247	99.9	245	2 C85916	hypothetical prote
3	1247	99.9	245	2 H91071	hypothetical prote
4	871.5	69.8	257	2 AF0396	probable amino aci
5	316.5	25.4	219	1 B69469	conserved hypothet
6	287	23.0	237	2 F84013	branched-chain ami
7	232	18.6	240	2 B97608	azlc family protei
8	232	18.6	240	2 AE2830	conserved hypothet
9	228	18.3	235	2 AB1255	transport proteins
10	225	18.0	235	2 AG1617	transport protein
11	212.5	17.0	235	2 G75494	Azlc family protei
12	209	16.7	252	2 H83390	hypothetical prote
13	208	16.7	239	2 G82388	Azlc family protei
14	204	16.3	235	2 E86826	amino acid permea
15	203.5	16.3	231	2 B89759	hypothetical prote
16	201	16.1	234	2 E97492	azlc family protei
17	201	16.1	234	2 AI2709	branched-chain ami
18	195.5	15.7	228	2 F71831	hypothetical prote
19	191.5	15.3	228	1 C64686	conserved hypothet
20	178	14.3	218	2 G95016	conserved hypothet
21	168	13.5	254	1 G69592	branched-chain ami
22	153	12.3	219	2 AE0945	probable exported
23	146	11.7	246	2 AD3574	branched-chain ami
24	145.5	11.7	256	2 T35133	hypothetical prote
25	136	10.9	244	1 F64041	hypothetical prote
26	135	10.8	275	2 D97525	hypothetical prote
27	124.5	10.0	160	2 C81877	hypothetical prote
28	109	8.7	1108	2 A48508	cyclic-nucleotide
29	107.5	8.6	397	2 G90048	hypothetical prote

30	104	8.3	651	2	S44257	phosphotransferase
31	103.5	8.3	300	2	E87280	rard protein [impo
32	103.5	8.3	410	2	AH0051	probable O-antigen
33	103.5	8.3	432	2	S67821	gumE protein - Xan
34	102	8.2	343	2	B69800	iron(III) dicitrat
35	101	8.1	467	2	E95850	probable amino aci
36	101	8.1	583	2	T12576	probable phosphate
37	100	8.0	406	2	A71109	probable phosphate
38	100	8.0	512	2	B70786	branched-integral
39	99.5	8.0	307	2	B36125	branched-chain ami
40	99	7.9	463	2	C69751	sodium/proton-depe
41	98.5	7.9	465	2	I39473	Na+-dependent phos
42	98.5	7.9	467	2	A48916	sodium phosphate t
43	98	7.9	429	2	B83826	hypothetical prote
44	97.5	7.8	460	2	A84154	amino acid transpo
45	97.5	7.8	517	2	AC2070	Na+/H+ antiporter

ALIGNMENTS

RESULT 1

C65048
hypothetical protein b2682 - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 05-Oct-2004
C;Accession: C65048
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: C65048
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-245 <BIAT>
A;Cross-references: UNIPROT:P76630; UNIPARC:UPI000004F5DD; GB:AE000353; GB:U00096; NID
A;Experimental source: strain K-12, substrain MG1655
C;Superfamily: branched-chain amino acid transport protein, Azlc type

Query Match 100.0%; Score 1248; DB 1; Length 245;
Best Local Similarity 100.0%; Pred. No. 3.8e-96;
Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MESPTPOPAPGSATFMEGCKDSLPIVISYIPVAFAGLNATRLGFSPLSFVFFSIIYAG	60
Db	1	MESPTPOPAPGSATFMEGCKDSLPIVISYIPVAFAGLNATRLGFSPLSFVFFSIIYAG	60
Qy	61	ASQFVITAMLAAGSSLIWIAALTVMAMDVRHVLVYGPSLRSRIIQRLQSKTALWAFGLTDE	120
Db	61	ASQFVITAMLAAGSSLIWIAALTVMAMDVRHVLVYGPSLRSRIIQRLQSKTALWAFGLTDE	120
Qy	121	VFAAATAKLVNRNRSENWMIAGFSSWSWVFTVIGAFSGSLGQGPVPAEALGFM	180
Db	121	VFAAATAKLVNRNRSENWMIAGFSSWSWVFTVIGAFSGSLGQGPVPAEALGFM	180
Qy	181	LPALFMSFLLASFORQSLCVTAALGALAGVTLFSPVAILAGIVCGCLTALIOAFWQOG	240
Db	181	LPALFMSFLLASFORQSLCVTAALGALAGVTLFSPVAILAGIVCGCLTALIOAFWQOG	240
Qy	241	APDEL 245	
Db	241	APDEL 245	

RESULT 2

C85916
hypothetical protein Z3983 [imported] - Escherichia coli (strain O157:H7, substrain BDL
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 05-Oct-2004
C;Accession: C85916
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca

Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: C85916
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-245 <STO>
A:Cross-references: UNIPROT:C8X908; UNIPROT:C8FEQ6; UNIPARC:UPI00000D0899; GB:AE005174;
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z3983
C:Superfamily: branched-chain amino acid transport protein, AzlC type

Query Match 99.9%; Score 1247; DB 2; Length 245;
Best Local Similarity 99.6%; Pred. No. 4.6e-96;
Matches 244; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MESPTQPAPGSATFMEGCKDSLPIVISYIPVAFAGLNATRLGFSPLSFVFFSCIIYAG 60
Db 1 MESPTQPAPGSATFMEGCKDSLPIVISYIPVAFAGLNATRLGFSPLSFVFFSCIIYAG 60

Qy 61 ASQFVITAMLAAGSSLWTAALTVMAMDVHVLVYGPSLSRSRIIQRLOKSKTALWAFGLTDE 120
Db 61 ASQFVITAMLAAGSSLWTAALTVMAMDVHVLVYGPSLSRSRIIQRLOKSKTALWAFGLTDE 120

Qy 121 VFAAATAKLVRNNRRWSNNWIGIAFSSWSWVFGTVIGAFSGSLGLOGYPVAAALGFM 180
Db 121 VFAAATAKLVRNNRRWSNNWIGIAFSSWSWVFGTVIGAFSGSLGLOGYPVAAALGFM 180

Qy 181 LPALFMSFLLASFORKQSLCVTAALVAGALAGVTLPFIPVAILAGIVCGCLTALIQAFWQ 240
Db 181 LPALFMSFLLASFORKQSLCVTAALVAGALAGVTLPFIPVAILAGIVCGCLTALIQAFWQ 240

Qy 241 APDEL 245
Db 241 APDEL 245

RESULT 3
H91071
hypothetical protein ECs3544 [imported] - Escherichia coli (strain O157:H7, substrain R1)
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 05-Oct-2004
C:Accession: H91071
R:Hayashi, T.; Makino, K.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; Sasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H. DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic reference number: A59629; MUID:21156231; PMID:11258796
A:Accession: H91071
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-245 <HAY>
A:Cross-references: UNIPROT:C8X908; UNIPROT:C8FEQ6; UNIPARC:UPI00000D0899; GB:BA0000007;
A:Experimental source: strain O157:H7, substrain R1MD 0509952
C:Genetics:
A:Gene: ECs3544
C:Superfamily: branched-chain amino acid transport protein, AzlC type

Query Match 99.9%; Score 1247; DB 2; Length 245;
Best Local Similarity 99.6%; Pred. No. 4.6e-96;
Matches 244; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MESPTQPAPGSATFMEGCKDSLPIVISYIPVAFAGLNATRLGFSPLSFVFFSCIIYAG 60
Db 1 MESPTQPAPGSATFMEGCKDSLPIVISYIPVAFAGLNATRLGFSPLSFVFFSCIIYAG 60

Qy 61 ASQFVITAMLAAGSSLWTAALTVMAMDVHVLVYGPSLSRSRIIQRLOKSKTALWAFGLTDE 120
Db 61 ASQFVITAMLAAGSSLWTAALTVMAMDVHVLVYGPSLSRSRIIQRLOKSKTALWAFGLTDE 120

Qy 121 VFAAATAKLVRNNRRWSNNWIGIAFSSWSWVFGTVIGAFSGSLGLOGYPVAAALGFM 180
Db 121 VFAAATAKLVRNNRRWSNNWIGIAFSSWSWVFGTVIGAFSGSLGLOGYPVAAALGFM 180

Db 121 VFAAATAKLVRNNRRWSNNWIGIAFSSWSWVFGTVIGAFSGSLGLOGYPVAAALGFM 180
Qy 181 LPALFMSFLLASFORKQSLCVTAALVAGALAGVTLPFIPVAILAGIVCGCLTALIQAFWQ 240
Db 181 LPALFMSFLLASFORKQSLCVTAALVAGALAGVTLPFIPVAILAGIVCGCLTALIQAFWQ 240
Qy 241 APDEL 245
Db 241 APDEL 245

RESULT 4
AF0396
probable amino acid transporter YPO3264 [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 05-Oct-2004
C:Accession: AF0396
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrrell, Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AF0396
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-257 <KUR>
A:Cross-references: UNIPROT:Q8ZBX0; UNIPARC:UPI00000DCA45; GB:AL590842; PIDN:CAC92498.1;
C:Genetics:
A:Gene: YPO3264
C:Superfamily: branched-chain amino acid transport protein, AzlC type

Query Match 69.8%; Score 871.5; DB 2; Length 257;
Best Local Similarity 70.9%; Pred. No. 5.9e-65;
Matches 168; Conservative 29; Mismatches 35; Indels 5; Gaps 1;

Qy 4 PTPQPAPGS-----ATFMEGCKDSLPIVISYIPVAFAGLNATRLGFSPLSFVFFSCIIY 58
Db 2 PNPTAFSSSTSAATFVEGITDSLPIVIGVLPVAFAGLSAVKLGFTPLSFVFFSCIIY 61

Qy 59 AGASQFVITAMLAAGSSLWTAALTVMAMDVHVLVYGPSLSRSRIIQRLOKSKTALWAFGLT 118
Db 62 AGASQFVITALLSAGMSLWVSALTVMAMDVHVLVYGPSLSRSRIIQRLOKSKTALWAFGLT 121

Qy 119 DEVFAAATAKLVRNNRRWSNNWIGIAFSSWSWVFGTVIGAFSGSLGLOGYPVAAALG 178
Db 122 DEVFAAATTKLMKQQRWSNNWMLGIAVTSWLSVWLGTAIGAMFGNGPLENPAIEASLS 181

Qy 179 FMLPALFMSFLLASFORKQSLCVTAALVAGALAGVTLPFIPVAILAGIVCGCLTALIQ 235
Db 182 FMLPALFMSFLLASFKRQYSLTVIASLGSALLGVLLPSIPVAILAGTAGGCLATLIQ 238

RESULT 5
B69469
conserved hypothetical protein AF1755 - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 05-Oct-2004
C:Accession: B69469
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L. Nature 390, 364-370, 1997
A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.; Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaea A:Reference number: A69250; MUID:98049343; PMID:9389475
A:Accession: B69469
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-219 <KLE>
A:Cross-references: UNIPROT:C28519; UNIPARC:UPI0000056BED; GB:AE000982; NIT
C:Superfamily: branched-chain amino acid transport protein, AzlC type

Query Match 25.4%; Score 316.5; DB 1; Length 219;
Best Local Similarity 35.8%; Pred. No. 4.4e-19;
Matches 82; Conservative 44; Mismatches 86; Indels 17; Gaps 6;

QY 12 SATMEGCKDSLPIVISYIPVAFAGLNATRLGFSPLSFVFCIIYAGASQFVITAMLA 71
DB 7 SAMFRKGLVYGFPIVMAIYIPVAFITGVLARTLGFSEVAMLASLLIFAGASQFALITLYS 66

QY 72 AGSSLWIAALTVMAMDVHVLVYGPSLRSRIIQRLOKSKTALMAFGLTDEVFAAATAKLVR 131
DB 67 --QSLSAIPFIPFLNLRHIYS---SIIAQLKLRPPHISAFGLTDEVFAVSV----- 115

QY 132 NNRWSENWMIATFSSWSWVFTGAFSGGLGQYPAVEAALGFMPLPALFMSFLLA 191
DB 116 -NSAENEFLLGLELGSYSWVGGLGALVLAGSTLILDRD-VYSALVFSISALFLVLLP 173

QY 192 SFORKQSLCVTAALVGLAGVTLSPVAILAGIVCGCLTALIOAFWOG 240
DB 174 NLKGRH---VRAAVSGGAVALA-FHLLNLTSVGIIAALAGPLLSGWDG 218

RESULT 6
F84013
branched-chain amino acid transporter BH2910 [imported] - Bacillus halodurans (strain C-
C/Species: Bacillus halodurans
C/Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 05-Oct-2004
C/Accession: F84013
R/Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A/Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A/Reference number: A83650; MUID:20512582; PMID:11058132
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-237 <STO>
A/Cross-references: UNIPROT:Q9K8U2; UNIPARC:UPI00000C4046; GB:AP001517; GB:BA0000004; NID
A/Experimental source: strain C-125
C/Genetics:
A/Gene: BH2910
C/Superfamily: branched-chain amino acid transport protein, AzLC type

Query Match 23.0%; Score 287; DB 2; Length 237;
Best Local Similarity 32.1%; Pred. No. 1.3e-16;
Matches 70; Conservative 49; Mismatches 87; Indels 12; Gaps 5;

QY 1 MESPTQPAPGSAFMEGCKDSLPIVISYIPVAFAGLNATRLGFSPLSFVFCIIYAG 60
DB 1 METASNMKARETDWLKGAIGSSIAIGYIPASLTFGLLAKGTGLTLETIAMSFLFYAG 60

QY 61 ASQFVITAMLAAGSSLMIAALTVMAMDVHVLVYGPSLRSRIIQRLOKSKTALMAFGLTDE 120
DB 61 ASQYMALTLIAIGTGTEIILTTFIVNIRHLLMSASIRAK-METHPVKRAITAFGITDE 119

QY 121 VFAAATAKLVRNRRWSENWMIATFSSWSWVFTGAFSGGLGQYPAVEAALGFMPLPALFMSFLLA 179
DB 120 VFALVTSQ---DRRLTNGFVIGAVIAYVSVVHSAVGYIVGSAL----PATLQQGMGV 171

QY 180 MLPALFMSFLLASFORQSLCVTAALVGLAGVTLFSI 217
DB 172 ALYAMFALLIPSVRKHRSVLILAGTAALLNG--LFSL 207

RESULT 7
B97608
azlc family protein [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C/Species: Agrobacterium tumefaciens
C/Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 05-Oct-2004
C/Accession: B97608
R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum

A/Reference number: A97359; MUID:21608551; PMID:11743194
A/Accession: B97608
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-240 <KUR>
A/Cross-references: UNIPROT:Q8UDQ3; UNIPARC:UPI00000D1D87; GB:AE007869; PIDN:AAK87819.1
C/Genetics:
A/Gene: AGR_C_3745
A/Map position: circular chromosome
C/Superfamily: branched-chain amino acid transport protein, AzLC type

Query Match 18.6%; Score 232; DB 2; Length 240;
Best Local Similarity 30.2%; Pred. No. 4.8e-12;
Matches 68; Conservative 37; Mismatches 112; Indels 8; Gaps 6;

QY 12 SATMEGCKDSLPIVISYIPVAFAGLNATRLGFSPLSFVFCIIYAGASQFVITAMLA 71
DB 3 NADFREGLKSGFPFIALSAAPFGALFGAVAVDNGLSITEATIMSGTVYAGASQLVGLF 62

QY 72 AGSSLWIAALTVMAMDVHVLVYGPSLRSRIIQRLOKSKTALMAFGLTDEVFAAATAKLVR 131
DB 63 QKVAPMLVLSVFAVNFPHILYSAAI-ARMISNWSLLQKAAGFFVLVDPQFAESVRK-YE 120

QY 132 NNRWSENWMIATFSSWSWVFTGAFSGGLGQYPAVEAALGFMPLPALFMSFLLA 191
DB 121 NTGTGFSWYMGFATPVYVILWMLAMTILGASLGN--LVGDPKA-IGLDVLLPIYFMGMVL- 176

QY 192 SFORKQSL--CVTAALVGLAGVTLSPVAILAGIVCGCLTALI 234
DB 177 SFRQNFYFVPLMASAGATVAFVGVSPHVSIGAVAGIVAVL 221

RESULT 8
AE2830
conserved hypothetical protein Atu2068 [imported] - Agrobacterium tumefaciens (strain C-
C/Species: Agrobacterium tumefaciens
C/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 05-Oct-2004
C/Accession: AE2830
R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo,
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kuttyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A/Reference number: AB2577; MUID:21608550; PMID:11743193
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-240 <KUR>
A/Cross-references: UNIPROT:Q8UDQ3; UNIPARC:UPI00000D1D87; GB:AE008688; PIDN:AAL43059.1
A/Experimental source: strain C58 (Dupont)
C/Genetics:
A/Gene: Atu2068
A/Map position: circular chromosome
C/Superfamily: branched-chain amino acid transport protein, AzLC type

Query Match 18.6%; Score 232; DB 2; Length 240;
Best Local Similarity 30.2%; Pred. No. 4.8e-12;
Matches 68; Conservative 37; Mismatches 112; Indels 8; Gaps 6;

QY 12 SATMEGCKDSLPIVISYIPVAFAGLNATRLGFSPLSFVFCIIYAGASQFVITAMLA 71
DB 3 NADFREGLKSGFPFIALSAAPFGALFGAVAVDNGLSITEATIMSGTVYAGASQLVGLF 62

QY 72 AGSSLWIAALTVMAMDVHVLVYGPSLRSRIIQRLOKSKTALMAFGLTDEVFAAATAKLVR 131
DB 63 QKVAPMLVLSVFAVNFPHILYSAAI-ARMISNWSLLQKAAGFFVLVDPQFAESVRK-YE 120

QY 132 NNRWSENWMIATFSSWSWVFTGAFSGGLGQYPAVEAALGFMPLPALFMSFLLA 191
DB 121 NTGTGFSWYMGFATPVYVILWMLAMTILGASLGN--LVGDPKA-IGLDVLLPIYFMGMVL- 176

C;Accession: H83390
 R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Boman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A;Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen
 A;Reference number: A82950; MUID:20437337; PMID:10984043
 A;Accession: H83390
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-252 <STO>
 A;Cross-references: UNIPROT:Q91276; UNIPARC:UPI00000C556B; GB:AE004630; GB:AE004091; NID
 A;Experimental source: strain PA01
 C;Genetics:
 A;Gene: PA2039
 C;Superfamily: branched-chain amino acid transport protein, AzLC type

Query Match 16.7%; Score 209; DB 2; Length 252;
 Best Local Similarity 28.3%; Pred. No. 4.1e-10;
 Matches 71; Conservative 39; Mismatches 109; Indels 32; Gaps 11;

Qy 5 TPQAPAGSATFMEGCKDSLPIVISYIPVAFAPGLNATRLGSPLESVFVFCIIYAGASQF 64
 Db 18 TAHDLPQA-FLRGAVAILPLSLAVAPWGLLAGSWATEANLTPAGGGLSAIVPAGAAQL 76

Qy 65 VITAMLAAGSLTAAITVWMDVRHVLVYGPSLSRRIQRLQSKTALWAFG-----LTDE 120
 Db 77 VAIGMLKGGAGFFSIITLLTLLTSHLLYGNLSRP-VLSPL----PGRWRIGLGLTLTDE 131

Qy 121 VFAATATKLVNRNRWENWMIIGTAFSSWSWVFGTIVGAFSGLLQGYPAVE-AALGF 179
 Db 132 FFALASQHRNFRNW---YALGGLTFYTAANLLFTLAILGRSI----PGLHGLDF 184

Qy 180 MLPALFMSFLLASFORKQSL-CVTAALVGLAGVTLFS-----IPVAILAGVCG--C 229
 Db 185 SIATFTALVAPLVNRVNTLVCVATSLFCV-----LFSHWQWSSALVLAGLWAGFIC 240

Qy 230 LTALIQAFWQG 240
 Db 241 NKLYREAPWSG 251

RESULT 13
 G82388
 AzLC family protein VCA1002 [imported] - *Vibrio cholerae* (strain N16961 serogroup O1)
 C;Species: *Vibrio cholerae*
 C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 05-Oct-2004
 C;Accession: G82388
 R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.; Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, B.L.; R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
 Nature 406, 477-483, 2000
 A;Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.
 A;Reference number: A82035; MUID:20406833; PMID:10952301
 A;Accession: G82388
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-239 <HEI>
 A;Cross-references: UNIPROT:Q9KKU8; UNIPARC:UPI00000C3712; GB:AE004427; GB:AE003853; NID
 A;Experimental source: serogroup O1; strain N16961; biotype El Tor
 C;Genetics:
 A;Gene: VCA1002
 A;Map position: 2
 C;Superfamily: branched-chain amino acid transport protein, AzLC type

Query Match 16.7%; Score 208; DB 2; Length 239;
 Best Local Similarity 28.9%; Pred. No. 4.7e-10;
 Matches 67; Conservative 42; Mismatches 101; Indels 22; Gaps 7;

Qy 2 ESPTQPAGSATFMEGCKDSLPIVISYIPVAFAPGLNATRLGSPLESVFVFCIIYAGA 61
 Db 10 DSPTP-----TRUFWQGTIAMLPLUSTIAVLPWGLLAGSFATEAGLSVIESQALSAYAGA 64

Qy 62 SQFVITAMLAAGSLTAAITVWMDVRHVLVYGPSLSRRIQRLQSKTALWAFGLTDEV 121
 Db 65 AQLVAIGMFTKAGLLSLIATPFIITSRHFLYSVMSRSKISPLPLRWRLTL-GFLLTDEL 123

Qy 122 FFAATATKLVNRNRWENWMIIGTAFSSWSWVFGTIVGAFSGLLQGYPAV-EAALGPM 180
 Db 124 FAICGAQSDKQFNW---YALGAGLSFYLINWLASLVGIVAGSYL----POLNQWGLEFA 176

Qy 181 LPALFMSFLLASFORKQSL-CVTAALVGLA-----AGVTLSFIPVAILAG 224
 Db 177 VAATFIAIVPNIKSWPVLISVLTALVSVLLTWGIEGSLMFASIGAMLAG 228

RESULT 14
 B86826
 amino acid permease yqfD [imported] - *Lactococcus lactis* subsp. *lactis* (strain IL1403)
 C;Species: *Lactococcus lactis* subsp. *lactis*
 C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 09-Jul-2004
 C;Accession: B86826
 R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrlich, G.
 Genome Res. 11, 731-753, 2001
 A;Title: The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp. *lactis* strain IL1403
 A;Reference number: A86625; MUID:21235186; PMID:11337471
 A;Accession: B86826
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-235 <STO>
 A;Cross-references: UNIPROT:Q9CF68; UNIPARC:UPI00000C6A96; GB:AE005176; PID:gl12724620;
 A;Experimental source: strain IL1403
 C;Genetics:
 A;Gene: yqfD

Query Match 16.3%; Score 204; DB 2; Length 235;
 Best Local Similarity 29.4%; Pred. No. 9.9e-10;
 Matches 59; Conservative 35; Mismatches 101; Indels 6; Gaps 3;

Qy 14 TFMGCKDSLPIVISYIPVAFAPGLNATRLGSPLESVFVFCIIYAGASQFVITAMLAAG 73
 Db 6 TFRQGLKDTPTVFGYIGIGIAFGMIGHSFGSVWILLSLIVYAGSAQFIMVSMLEATH 65

Qy 74 SSLWIAALTVWMDVRHVLVYGPSLSRRIQRLQSKTALWAFGLTDEVFAAATAKLVRNN 133
 Db 66 SPMSIVLSVFLVNSRIILMSMTTAS-YFKNESLLKNILGTLTLLTDSFALGMNKQNYTE 124

Qy 134 RRSENWMIIGTAFSSWSWVFGTIVGAFSGLLQGYPAVEAALGFMPLPALSFLAS 193
 Db 125 GKLNFSWFRNSNLLAVLWALASAGALLGNLLAN---PEKLGFGFAVIAFMFGLLYQL 181

Qy 194 QRKQSLCVTAALVGLAGVTL 214
 Db 182 ISDKTLGLMLQLV--MVGITL 200

RESULT 15
 B89759
 hypothetical protein SA0010 [imported] - *Staphylococcus aureus* (strain N315)
 C;Species: *Staphylococcus aureus*
 C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
 C;Accession: B89759
 R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogu, M.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogatawara, N.; Hayashi, H.; Hiramatsu, K.
 Lancet 357, 1225-1240, 2001
 A;Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.
 A;Reference number: A89758; MUID:21311952; PMID:11418146
 A;Accession: B89759
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-231 <KUR>
 A;Cross-references: UNIPROT:Q9XKG1; UNIPARC:UPI00000CAAL13; GB:BA000018; PID:gl13699927;
 A;Experimental source: strain N315
 C;Genetics:
 A;Gene: SA0010

Protein Sequence Searches - February 2005

All of the sequence databases on ABSS have recently been updated.

- Please note that the curators of the UniProt database have purged some temporary accession numbers from the most recent version of UniProt. These sequences have been assigned new permanent accession numbers. The new UniProt record may not contain the previous temporary accession number.
- If you encounter an accession number from an older search run against UniProt (results file extension **.rup**) that can no longer be found in the database, the permanent record with the new accession number can be found by searching the old accession number in the UniProt Protein Archive database (UniPARC) at:

<http://www.pir.uniprot.org/database/archive.shtml>

If you have any questions regarding this information or your results, please contact any STIC searcher.

When submitting sequence search results for scanning into IFW, please include a copy of this attachment to assist any future Examiners or members of the public who may encounter UniProt temporary accession numbers.

This Page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 12:16:07 ; Search time 203.02 Seconds
(without alignments)
851.417 Million cell updates/sec

Title: US-10-073-293A-4
Perfect score: 1348
Sequence: 1 MESPTQPAPGSATFMEGCK.....VCGCLTALIOAFMQGAPDEL 245

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot_05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1248	100.0	245	1 YGAZ_ECOLI	P76630 escherichia
2	1247	99.9	245	2 Q8FEQ6_ECOL6	Q8FEQ6 escherichia
3	1247	99.9	245	2 Q8X908_ECO57	Q8X908 escherichia
4	1243	99.6	245	2 Q83J26_SHIFL	Q83J26 shigella fl
5	871.5	69.8	257	2 Q8ZBX0_YERPE	Q8ZBX0 yersinia pe
6	870.5	69.8	257	2 Q6EE34_YERFS	Q6EE34 yersinia fs
7	845.5	67.7	243	2 Q6DID9_ERWCT	Q6DID9 erwinia car
8	841.5	67.4	252	2 Q7N779_PHOLL	Q7N779 photorhabdu
9	834	66.8	247	2 Q93KB4_ERWCH	Q93KB4 erwinia chr
10	385	30.8	237	2 Q7WB32_BORPA	Q7WB32 bordetella
11	385	30.8	237	2 Q7WMJ8_BORBR	Q7WMJ8 bordetella
12	316.5	25.4	219	1 Y1755_ARCFU	Y1755 archaeoglob
13	308.5	24.7	224	2 Q8VW59_BRUBA	Q8VW59 brucella ab
14	308.5	24.7	224	2 Q8FY15_BRUSA	Q8FY15 brucella su
15	301.5	24.2	238	2 Q6LZP2_METWP	Q6LZP2 methanococ
16	287	23.0	237	2 Q9K8U2_BACHD	Q9K8U2 bacillus ha
17	286.5	23.0	216	2 Q5SLQ8_THET8	Q5SLQ8 thermus the
18	286.5	23.0	216	2 Q72GU7_THET2	Q72GU7 thermus the
19	281	22.5	238	2 Q5L272_GEOKA	Q5L272 geobacillus
20	276	22.1	234	2 Q67SA1_SYMTH	Q67SA1 symbiobacte
21	269	21.6	241	2 Q6HKN2_BACHK	Q6HKN2 bacillus th
22	269	21.6	241	2 Q81SD8_BACAN	Q81SD8 bacillus li
23	268.5	21.5	238	2 Q65LQ0_BACLD	Q65LQ0 bacillus li
24	268	21.5	241	2 Q63D72_BACIL	Q63D72 bacillus ce
25	267	21.4	241	2 Q4MUB4_BACCE	Q4MUB4 bacillus ce
26	267	21.3	241	2 Q73AI3_BACCI	Q73AI3 bacillus ce
27	262	21.0	241	2 Q81FC0_BACCR	Q81FC0 bacillus ce
28	254	20.4	242	2 Q73IS3_BACCI	Q73IS3 bacillus ce
29	251	20.1	233	2 Q6AKD1_DESFS	Q6AKD1 desulfocale
30	248	19.9	243	2 Q5WAZ4_BACSK	Q5WAZ4 bacillus cl
31	247.5	19.8	242	2 Q8Y223_RALSON	Q8Y223 ralstonia s

32	243	19.5	243	2	Q4FS78_9GAMM	Q4fs78 psychrobact
33	240.5	19.3	235	2	Q83IJ7_ENTFA	Q83ij7 enterococcu
34	238	19.1	234	2	Q74F41_GEOSL	Q74f41 geobacter s
35	236	18.9	242	2	Q98KC6_RHILO	Q98kc6 rhizobium l
36	236	18.9	243	2	Q4NSK9_9DELT	Q4nsk9 anaeromyxob
37	235.5	18.9	230	2	Q8EG49_SHEON	Q8eg49 shewanella
38	235	18.8	232	2	Q4FL99_9RICK	Q4fl99 candidatus
39	234	18.8	232	2	Q5LY33_STRT1	Q5ly33 streptococc
40	234	18.8	232	2	Q5M2N8_STRT2	Q5m2n8 streptococc
41	232	18.6	240	2	Q8UDQ3_AGRTS	Q8udq3 agrobacteri
42	231.5	18.5	232	2	Q8DVS9_STRMU	Q8dvs9 streptococc
43	228.5	18.3	241	2	Q8CU18_OCEIH	Q8cul8 oceanobacil
44	228	18.3	230	2	Q5V111_HALMA	Q5v111 haloarcula
45	228	18.3	235	2	Q8Y761_LISMO	Q8y761 listeria mo

ALIGNMENTS

RESULT 1

ID	YGAZ_ECOLI	STANDARD	PRT	245 AA.
AC	P76630			
DT	15-JUL-1998	(Rel. 36, Created)		
DT	15-JUL-1998	(Rel. 36, Last sequence update)		
DT	13-SEP-2005	(Rel. 48, Last annotation update)		
DE	Inner membrane protein ygaZ.			
GN	Names:ygaZ; OrderedLocusNames=b2682;			
OS	Escherichia coli.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Escherichia.			
OX	NCBI_TaxID=562;			
RN	[1]			
RP	NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].			
RC	STRAIN=K12 / MG1655;			
RC	MDLLINE=9746617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;			
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,			
RA	Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.P.,			
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,			
RA	Mau B., Shao Y.;			
RL	"The complete genome sequence of Escherichia coli K-12.";			
RL	Science 277:1453-1474 (1997).			
RN	[2]			
RP	TOPOLOGY.			
RC	STRAIN=K12 / MG1655;			
RC	PubMed=15919996; DOI=10.1126/science.1109730;			
RA	Daley D.O., Rapp M., Graneth E., Melen K., Drew D., von Heijne G.;			
RT	"Global topology analysis of the Escherichia coli inner membrane proteome.";			
RL	Science 308:1321-1323(2005).			
CC	-I- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.			
CC	-I- SIMILARITY: Belongs to the azlC family.			
CC	This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.			
CC	EMBL; U00096; AAC75729.1; -; Genomic_DNA.			
DR	PIR; C65048; C65048.			
DR	EcoBASE; EB3299; -.			
DR	EcoGene; EG13528; YgaZ.			
DR	InterPro; IPR011606; AzlC like.			
DR	InterPro; IPR012294; TFIID_C(glycos_N.			
DR	Pfam; PF03591; AzlC; 1.			
DR	Complete proteome; Inner membrane; Membrane; Transmembrane; Transport.			
FT	TOPO_DOM 1 24 Cytoplasmic (Potential).			
FT	TRANSMEM 25 45 Potential.			
FT	TOPO_DOM 64 63 Periplasmic (Potential).			
FT	TRANSMEM 46 84 Potential.			
FT	TOPO_DOM 85 109 Cytoplasmic (Potential).			
FT	TRANSMEM 110 130 Potential.			

```

FT TOPO DOM 131 140 Periplasmic (Potential).
FT TRANSEM 141 161 Potential.
FT TOPO_DOM 162 172 Cytoplasmic (Potential).
FT TRANSEM 173 193 Potential.
FT TOPO_DOM 194 205 Periplasmic (Potential).
FT TRANSEM 206 226 Potential.
FT TOPO_DOM 227 245 Cytoplasmic (Potential).
SQ SEQUENCE 245 AA; 26108 MW; 22AC8AB8D7D651B CRC64;

Query Match 100.0%; Score 1248; DB 1; Length 245;
Best Local Similarity 100.0%; Pred. No. 7.e-93;
Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MESPTQPAPGSATFMEGCKDSLPIVISYIPVAFGLNATRLGFSPLSFVFFSCIYAG 60
Db 1 MESPTQPAPGSATFMEGCKDSLPIVISYIPVAFGLNATRLGFSPLSFVFFSCIYAG 60
Qy 61 ASQFVITAMLAAGSSLWIAALTVMAMDVRHLYGFSLSRSRIIQRLOKSKTALWAFGLTDE 120
Db 61 ASQFVITAMLAAGSSLWIAALTVMAMDVRHLYGFSLSRSRIIQRLOKSKTALWAFGLTDE 120
Qy 121 VFAAATAKLVRNRRNSNMWIGIAFSSWSWVFTGTVIGAFSGSLGQGYPAVEAALGFM 180
Db 121 VFAAATAKLVRNRRNSNMWIGIAFSSWSWVFTGTVIGAFSGSLGQGYPAVEAALGFM 180
Qy 181 LPALFMSFLLASFORQKSLCVTAALVGLAGVTLFSPVAILAGIVCGCLTALIQAFWQG 240
Db 181 LPALFMSFLLASFORQKSLCVTAALVGLAGVTLFSPVAILAGIVCGCLTALIQAFWQG 240
Qy 241 APDEL 245
Db 241 APDEL 245

RESULT 3
Q8X908 ECO57 PRELIMINARY; PRT; 245 AA.
AC Q8X908; Q7ABD6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Hypothetical protein ECs3544.
GN OrderedLocusNames=ECs3544, z3983;
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=O157:H7 / EDL933 / ATCC 700927 / EHEC;
RX MEDLINE=21074935; PubMed=11206551; DOI=10.1038/35054089;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glaesner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamocous K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RL Nature 409:529-533(2001).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=O157:H7 / Sakai / RIMD 050952 / EHEC;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayaishi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
DR EMBL; AE005174; AAG57791.1; -; Genomic_DNA.
DR EMBL; BA000007; BAB36967.1; -; Genomic_DNA.
DR PIR; C85916; C85916.
DR PIR; H91071; H91071.
DR GO; GO:0003677; F:DNA binding; IEA.
DR InterPro; IPR011606; AzIC like.
DR InterPro; IPR012294; TFIID_C/glycos_N.
DR Pfam; PF03591; AzIC; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 245 AA; 26093 MW; 20AAF2D91E644DB3 CRC64;

Query Match 99.9%; Score 1247; DB 2; Length 245;
Best Local Similarity 99.6%; Pred. No. 9.2e-93;
Matches 244; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MESPTQPAPGSATFMEGCKDSLPIVISYIPVAFGLNATRLGFSPLSFVFFSCIYAG 60

```


Db 1 MESPTQPAPGATFMGCKDSLPIVISIPVAFAGLNATRLGSPLESVFFSCIIYAG 60
 QY 61 ASQFVITAMLAAGSSLIWIAALTVMAMDVHRHLYGSPSLRSRIIQRLOKSKTALWAFGLTDE 120
 Db 61 ASQFVITAMLAAGSSLIWIAALTVMAMDVHRHLYGSPSLRSRIIQRLOKSKTALWAFGLTDE 120
 QY 121 VFAAATAKLVRNRRWSNWMIGIAFSSWSWVFQTVIGAFSGSGLLGYPAVEAALGFM 180
 Db 121 VFAAATAKLVRNRRWSNWMIGIAFSSWSWVFQTVIGAFSGSGLLGYPAVEAALGFM 180
 QY 181 LPALPMSFLLASFORKQSLCVTAALVGLAGVTLFSPVAILAGIVCCCLTALIOAFWOG 240
 Db 181 LPALPMSFLLASFORKQSLCVTAALVGLAGVTLFSPVAILAGIVCCCLTALIOAFWOG 240
 QY 241 APDEL 245
 Db 241 APDEL 245

RESULT 4

Q83JZ6_SHIFL
 ID Q83JZ6 SHIFL PRELIMINARY; PRT; 245 AA.
 AC Q83JZ6; Q7C087;
 DT 01-JUN-2003 (TRENBLrel. 24, Created)
 DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
 DT 13-SEP-2005 (TRENBLrel. 31, Last annotation update)
 DE Hypothetical protein.
 GN OrderedLocusNames=S2896, SP2709;
 OS Shigella flexneri.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Shigella.
 OX NCBI_TaxID=623;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=301 / Serotype 2a;
 RA MEDLINE=2227406; PubMed=12384590; DOI=10.1093/nar/gkf566;
 RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
 RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
 RA Sun L., Xue Y., Zhao X., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
 RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
 RA Yu J.;
 RT "Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity
 RT through comparison with genomes of *Escherichia coli* K12 and O157.";
 RL Nucleic Acids Res. 30:4432-4441(2002).
 RN [2]

NUCLEOTIDE SEQUENCE

RC STRAIN=2457T / ATCC 700930 / Serotype 2a;
 RX MEDLINE=22590274; PubMed=12704152;
 RX DOI=10.1128/JAI.71.5.2775-2786.2003;
 RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
 RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
 RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
 RA Schwartz D.C., Blattner F.R.;
 RT "Complete genome sequence and comparative genomics of *Shigella*
 RT flexneri serotype 2a strain 2457T.";
 RL Infect. Immun. 71:2775-2786(2003).
 RN [3]
 DR ENBL; AE005674; AAN44202.1; -; Genomic DNA.
 DR ENBL; AE016987; AAP18029.1; -; Genomic_DNA.
 DR GO; GO:0003677; P:DNA binding; IEA.
 DR InterPro; IPRO11606; AzIC like.
 DR InterPro; IPRO12294; TFIID C/glycos_N.
 DR Pfam; PF03591; AzIC; 1.
 KW Complete proteome; Hypothetical protein.
 SQ SEQUENCE 245 AA; 26079 MW; 3BAC440092FB67F6 CRC64;

Query Match 99.6%; Score 1243; DB 2; Length 245;
 Best Local Similarity 99.2%; Pred. No. 1.9e-92;
 Matches 243; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MESPTQPAPGATFMGCKDSLPIVISIPVAFAGLNATRLGSPLESVFFSCIIYAG 60
 Db 1 MESPTQPAPGATFMGCKDSLPIVISIPVAFAGLNATRLGSPLESVFFSCIIYAG 60

RESULT 5

Q8ZBX0_YERPE
 ID Q8ZBX0 YERPE PRELIMINARY; PRT; 257 AA.
 AC Q8ZBX0; Q74W29; Q7CK78;
 DT 01-MAR-2002 (TRENBLrel. 20, Created)
 DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
 DT 01-FEB-2005 (TRENBLrel. 29, Last annotation update)
 DE Putative amino acid transporter (Hypothetical protein y0925).
 GN Name=azlC; OrderedLocusNames=YPO668, YPO3264, Y0925;
 OS *Yersinia pestis*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Yersinia*.
 OX NCBI_TaxID=632;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=CO-92 / Biovar Orientalis;
 RA MEDLINE=21470413; PubMed=11586360; DOI=10.1038/35097083;
 RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
 RA Prentice M.B., Sebahia M., James K.D., Churcher C.M., Mungall K.L.,
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.-M.,
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
 RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
 RA Leather S., Moulé S., Oyston P.C.F., Quail M.A., Rutherford K.M.,
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;
 RT "Genome sequence of *Yersinia pestis*, the causative agent of plague.";
 RL Nature 413:523-527(2001).
 RN [2]

NUCLEOTIDE SEQUENCE

RC STRAIN=KIM5 / Biovar Mediaevalis;
 RX MEDLINE=22137863; PubMed=12142430;
 RX DOI=10.1128/JB.184.16.4601-4611.2002;
 RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
 RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
 RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
 RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
 RA Perry R.D.;
 RT "Genome sequence of *Yersinia pestis* KIM.";
 RL J. Bacteriol. 184:4601-4611(2002).
 RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=91001;
 RX PubMed=15368893;
 RA Song Y., Tong Z., Wang J., Wang L., Guo Z., Han Y., Zhang J., Pei D.,
 RA Zhou D., Qin H., Pang X., Han Y., Zhai J., Li M., Cui B., Qi Z.,
 RA Jin L., Dai R., Chen F., Li S., Ye C., Du Z., Lin W., Wang J., Yu J.,
 RA Yang H., Wang J., Huang P., Yang R.;
 RT "Complete genome sequence of *Yersinia pestis* strain 91001, an isolate
 RT avirulent to humans.";
 RL DNA Res. 11:179-197(2004).
 DR ENBL; AJ414156; CAC92498.1; -; Genomic DNA.
 DR ENBL; AE013695; AAM84507.1; -; Genomic_DNA.
 DR ENBL; AE017129; AAS60935.1; -; Genomic_DNA.
 DR PIR; AF0396; AF0396.
 DR InterPro; IPRO11606; AzIC like.
 DR Pfam; PF03591; AzIC; 1.
 KW Complete proteome; Hypothetical protein.

```
SQ SEQUENCE 257 AA; 27301 MW; 7C3660B0C8AD4FFPB CRC64;

Query Match 69.8%; Score 871.5; DB 2; Length 257;
Best Local Similarity 70.9%; Pred. No. 2.1e-62;
Matches 168; Conservative 29; Mismatches 35; Indels 5; Gaps 1;

QY 4 PTPQAPGS-----ATFMEGCKDSLPIVISYIPVAFAGLNATRLGSPLESVFFSCIIY 58
DB 2 ENPTAPSTTSAAATFVEGTDLSLPVIGVLPVAFAGLSAVKLGFTPLSFISFSCIIY 61

QY 59 AGAQFVITAMLAAGSSLIWIAALTMVMDVRHLYGSPLSRSRIIRLQKSKTALWAFGLT 118
DB 62 AGAQFVITALLSAGMSLWVSALTVMMDVRHLYGPAKHLIRLAKLGGKKTALWAFGLT 121

QY 119 DEVFAAATAKLVRNRRSENWMIQAFSSSSWVFGTVICAFSGSLGQYPAVEAALG 178
DB 122 DEVFAAATTKLMKQRRSENWMLGIAVTLSWLVGTAIGAMFGNGPLENYPALAEASLS 181

QY 179 FMLPALFMSFLASFKRQKSLVTIASLSGALLGVLLFSIPVAILAGIAGGCLATLIQ 235
DB 182 FMLPALFMSFLASFKRQKSLVTIASLSGALLGVLLFSIPVAILAGIAGGCLATLIQ 238

RESULT 6
Q66E34_YERPS PRELIMINARY; PRT; 257 AA.
AC Q66E34_YERPS PRELIMINARY; PRT; 257 AA.
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DE Putative LIV-E family branched chain amino acid exporter, large subunit.
GN Name=YgagZ; OrderedLocNames=YPTB0859;
OS Yersinia pseudotuberculosis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=633;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=IP32953 / Serotype 1;
RX PubMed=15358958; DOI=10.1073/pnas.0404012101;
RA Chain P.S.G., Carniel E., Larimer F.W., Lamerdin J., Stoutland P.O., Regala W.M., Georgescu A.M., Vergez L.M., Land M.L., Motin V.L., Brubaker R.R., Fowler J., Hinebusch J., Marceau M., Medigue C., Simonet M., Chenail-Francisque V., Souza B., Dacheux D., Elliott J.M., Dextise A., Hauser L.J., Garcia E.;
RA "Insights into the evolution of Yersinia pestis through whole-genome comparison with Yersinia pseudotuberculosis.";
RT Proc. Natl. Acad. Sci. U.S.A. 101:13826-13831(2004).
RL EMBL; BX936398; CAH2009.1; -; Genomic_DNA.
DR InterPro; IPR011606; AzlC; 1.
DR Pfam; PF03591; AzlC; 1.
KW Complete proteome.
SQ SEQUENCE 257 AA; 27297 MW; 88078AFD3104E45A CRC64;

Query Match 69.8%; Score 870.5; DB 2; Length 257;
Best Local Similarity 70.9%; Pred. No. 2.6e-62;
Matches 168; Conservative 29; Mismatches 35; Indels 5; Gaps 1;

QY 4 PTPQAPGS-----ATFMEGCKDSLPIVISYIPVAFAGLNATRLGSPLESVFFSCIIY 58
DB 2 ENPTAPSTTSAAATFVEGTDLSLPVIGVLPVAFAGLSAVKLGFTPLSFISFSCIIY 61

QY 59 AGAQFVITAMLAAGSSLIWIAALTMVMDVRHLYGSPLSRSRIIRLQKSKTALWAFGLT 118
DB 62 AGAQFVITALLSAGMSLWVSALTVMMDVRHLYGPAKHLIRLAKLGGKKTALWAFGLT 121

QY 119 DEVFAAATAKLVRNRRSENWMIQAFSSSSWVFGTVICAFSGSLGQYPAVEAALG 178
DB 122 DEVFAAATTKLMKQRRSENWMLGIAVTLSWLVGTAIGAMFGNGPLENYPALAEASLS 181

QY 179 FMLPALFMSFLASFKRQKSLVTIASLSGALLGVLLFSIPVAILAGIAGGCLATLIQ 235
DB 182 FMLPALFMSFLASFKRQKSLVTIASLSGALLGVLLFSIPVAILAGIAGGCLATLIQ 238
```

```
DB 182 FMLPALFMSFLASFKRQKSLVTIASLSGALLGVLLFSIPVAILAGIAGGCLATLIQ 238

RESULT 7
Q6D1D9_ERWCT PRELIMINARY; PRT; 243 AA.
AC Q6D1D9_ERWCT PRELIMINARY; PRT; 243 AA.
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DE Putative amino acid transporter.
GN OrderedLocNames=ECA3509;
OS Erwinia carotovora (subsp. atroseptica) (Pectobacterium atrosepticum).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pectobacterium.
OX NCBI_TaxID=29471;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=SCRI 1043 / ATCC BAA-672;
RX PubMed=15263089; DOI=10.1073/pnas.0402424101;
RA Bell K.S., Sebahnia M., Pritchard L., Holden M.T.G., Hyman L.J., Holeva M.C., Thomson N.R., Brooks K., Chillingworth T., Clark K., Doggett J., Atkin R., Bason N., Hauser H., Jagels K., Moule S., Norbertczak H., Fraser A., Hance Z., Quail M.A., Sanders M., Walker D., Whitehead S., Ormond D., Price C., Birch P.R.J., Parkhill J., Toth I.K.;
RA "Genome sequence of the enterobacterial phytopathogen Erwinia carotovora subsp. atroseptica and characterization of virulence factors.";
RT Proc. Natl. Acad. Sci. U.S.A. 101:11105-11110(2004).
RL EMBL; BX950851; CAG76407.1; -; Genomic_DNA.
DR InterPro; IPR011606; AzlC_like.
DR Pfam; PF03591; AzlC; 1.
KW Complete proteome.
SQ SEQUENCE 243 AA; 25805 MW; B00PB8F1B40A2EF1 CRC64;

Query Match 67.7%; Score 845.5; DB 2; Length 243;
Best Local Similarity 70.3%; Pred. No. 2.6e-60;
Matches 163; Conservative 28; Mismatches 40; Indels 1; Gaps 1;

QY 5 TPQAPGSATFMEGCKDSLPIVISYIPVAFAGLNATRLGSPLESVFFSCIIYAGASQF 64
DB 6 TPAPTK-SASFREGVDSLPVIGVMPVAFAGFNVAVKLGFTPLGIFLSCIIYAGASQF 64

QY 65 VITAMLAGSSLIWIAALTMVMDVRHLYGSPLSRSRIIRLQKSKTALWAFGLTDEVPA 124
DB 65 VITALLSAGMSIWWAALTVMMDIRHLYGPAKHLIRVQQPTTKTALWAFGLTDEVPA 124

QY 125 ATAKLVRRNRRSENWMIQAFSSSSWVFGTVICAFSGSLGQYPAVEAALGFMLPAL 184
DB 125 AATRLAKDNRNSENWMIQAFSSSSWVFGTVICAFSGSLGQYPAVEAALGFMLPAL 184

QY 185 FMSFLASFKRQKSLVTIASLSGALLGVLLFSIPVAILAGIAGGCLATLIQ 236
DB 185 FLSFLASFKRQKSLVTIASLSGALLGVLLFSIPVAILAGIAGGCLATLIQ 236

RESULT 8
Q7N779_PHOLL PRELIMINARY; PRT; 252 AA.
AC Q7N779_PHOLL PRELIMINARY; PRT; 252 AA.
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DE Similar to unknown protein Ygag of Escherichia coli.
GN OrderedLocNames=plu1279;
OS Photobacterium luminescens (subsp. laumondii).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photobacterium.
OX NCBI_TaxID=141679;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=TT01;
```

RESULT 1C

Q7WB32 BORPA
ID Q7WB32 BORPA PRELIMINARY; PRT; 237 AA.

Qv 1 MESPTROBAPCSATEMEGCKNSIPIVITSYIPVAFEGINATPIGESPIESIVERECTIYAC 60

Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	Q21	Q22	Q23	Q24	Q25	Q26	Q27	Q28	Q29	Q30	Q31	Q32	Q33	Q34	Q35	Q36	Q37	Q38	Q39	Q40	Q41	Q42	Q43	Q44	Q45	Q46	Q47	Q48	Q49	Q50	Q51	Q52	Q53	Q54	Q55	Q56	Q57	Q58	Q59	Q60	Q61	Q62	Q63	Q64	Q65	Q66	Q67	Q68	Q69	Q70	Q71	Q72	Q73	Q74	Q75	Q76	Q77	Q78	Q79	Q80	Q81	Q82	Q83	Q84	Q85	Q86	Q87	Q88	Q89	Q90	Q91	Q92	Q93	Q94	Q95	Q96	Q97	Q98	Q99	Q100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100

Db 6 IDAPGPPGAPVSA---GLKACVPVMIGYFPVAVAFGIAGLAAGLQPLQVILISVFVYAG 61

Q. 61 ACCIDENTALITY ACCORDS - I WANT TO EXAMINE THE VARIOUS DEPT. OF CORRECTIONS MAINTENANCE

QY 01 ASQV I I A M L A G S S - D W I A H A L I V N A H D V K H V L I G F S L R S K I I Q K L Q S K I A L W A F G L I D I I S

Db 62 ASQFLLASIKAGTPWLWVALCSL-LNARHLLYGP-LLARFLPESLRERLRI-AFLITD 118

[illegible]

QY 120 EVFAAATAKLVNRNRRWSENMMIGIAFSSWSSWVFGIVIGAFSGSGLLQGYPFAVEAALGF 179

Db 119 EVFATAFNRIGAVEPASRGRWITALGLGAWATWIAGTAVGVYAGEGLERHYPVLSQVMRF 178

[illegible]

180 MLPALEMSFLLASFQKQSLCVTAALV--GALAGVTLFSIPVAILAGIVCGCL 230

Db 179 ALPALFLALVCQSMQPGMRRPVLAAALAVGGALAAVG--QTTLAILAGAAAGCL 229

PGSUT 11

RESOLUT 11
07WMJ8 BORBR

ID Q7WMJ8 BORBR PRELIMINARY; PRT; 237 AA.

AC Q7WMJ8;
DE 01 OCT 2003 /E-EMBT-01 25 C-000001

DI	01-OCT-2003 (TREMBLER: 23, Created)
DT	01-OCT-2003 (TREMBLER: 25, Last sequence update)

DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)

DE	Putative membrane protein.	RA	Zhou L., Overbeek R., Gocayne J.D., Weidman J.F., McDonald L.A.,
GN	OrderedLocusNames=BB1393;	RA	Uterback T.R., Cotton M.D., Spriggs T., Artlich P., Kaine B.P.,
GN	Bordetella bronchiseptica (Alcaligenes bronchisepticus).	RA	Sykes S.M., Sadow P.W., D'Andrea K.P., Bowman C., Fujii C.,
OC	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;	RA	Garland S.A., Mason T.M., Olsen G.J., Fraser C.M., Smith H.O.,
OC	Alcaligenaceae; Bordetella.	RA	Woese C.R., Venter J.C.;
OX	NCBI_TaxID=518;	RT	"The complete genome sequence of the hyperthermophilic, sulphate-
NC	[1]	RT	reducing archaeon Archaeoglobus fulgidus."
RP	NUCLEOTIDE SEQUENCE.	RL	Nature 390:364-370(1997).
RC	STRAIN=RB50 / ATCC BAA-588;	CC	!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
RC	MEDLINE=2827954; PubMed=12910271; DOI=10.1038/ng1227;	CC	!- SIMILARITY: Belongs to the azlC family.
RA	Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.R.,	CC	
RA	Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,	CC	This Swiss-Prot entry is copyright. It is produced through a collaboration
RA	Cerdeno-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,	CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
RA	Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,	CC	the European Bioinformatics Institute. There are no restrictions on its
RA	Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,	CC	use as long as its content is in no way modified and this statement is not
RA	Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,	CC	removed.
RA	Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,	CC	
RA	Rabinowitz E., Rutter S., Sanders M., Saunders D., Seeger K.,	CC	
RA	Sharp S., Simmonds M., Skellern J., Squares R., Squares S., Stevens K.,	CC	
RA	Unwin L., Whitehead S., Barrall B.G., Maskell D.J.;	CC	
RT	"Comparative analysis of the genome sequences of Bordetella pertussis,	CC	
RT	Bordetella parapertussis and Bordetella bronchiseptica.;"	CC	
RL	Nat. Genet. 35:32-40(2003).	CC	
DR	EMBL; BX640441; CAE31891.1; -; Genomic_DNA.	DR	EMBL; AE000982; AAB89495.1; -; Genomic_DNA.
DR	InterPro; IPR011606; AzlC_like.	DR	PIR; B69469; B69469.
DR	pfam; PF03591; AzlC; 1.	DR	TIGR; AF1755; -.
KW	Complete proteome.	DR	InterPro; IPR011606; AzlC_like.
KW	Complete proteome.	DR	InterPro; IPR011606; AzlC; 1.
SQ	SEQUENCE 237 AA; 24835 MW; EF448ED5E37F54E0 CRC64;	KW	Complete proteome; Hypothetical protein; Transmembrane; Transport.
		FT	TRANSMEM 14 34 Potential.
		FT	TRANSMEM 37 57 Potential.
		FT	TRANSMEM 123 143 Potential.
		FT	TRANSMEM 155 175 Potential.
		FT	TRANSMEM 189 209 Potential.
		SQ	SEQUENCE 219 AA; 23338 MW; 265A2C88DA93EB86 CRC64;
		Query Match	25.4%; Score 316.5; DB 1; Length 219;
		Best Local Similarity	35.8%; Pred. No. 1.3e-17;
		Matches	82; Conservative 44; Mismatches 86; Indels 17; Gaps 6;
QY	1 MSPTPQPAPGATMEGCKSLPIVSIYPVAFAGLNATLGFSPLESVPFSCIYAG 60	QY	12 SATFMGCKSLPIVSIYPVAFAGLNATLGFSPLESVPFSCIYAGASQFVITAMLA 71
DB	6 IDAPGPGAPVSA---GLKACVPVMIGYFPVAFAGIAGLQPLQVILISVFYAG 61	DB	7 SAMFRKGIYVFPVIMAYIPVAFATLGFSEVEAMLSLIPAGASQFALITLYS 66
QY	61 ASQFVTAMLAAGSS-LWTAALTVMADVHRVLYGPSLSRIIQRLQKSKTALWAFGLTD 119	QY	72 AGSSLIWIALTVMADVHRVLYGPSLSRIIQRLQKSKTALWAFGLTDEFAATAKLVR 131
DB	62 ASQFLLASIKAGTQPLWVVALCSF-LNARHLVGP-LIARPLPESRLRI-AFLTLD 118	DB	67 --QSLSAIFPIFLNLRHIYS---SITAQKLRFPHISAFGLTDEFAVSVV---- 115
QY	120 EVFAAATAKLVRNRNRWSNMWIGAFSSWSVFGTVIGAFSGSLGQGYPAVEAALGF 179	QY	132 NNRRWSNMWIGAFSSWSVFGTVIGAFSGSLGQGYPAVEAALGFMLPALFMSFLA 191
DB	119 EVFATAFNRIGAVEPASRGWITLGLGAWATWIAGTAVGYAGEGLEHYPLVSQVRF 178	DB	116 -NSAENERFLGLEIGSYSVMVGTALGVLAGSTLILDRD-VYSALVFSISALFLVLLP 173
QY	180 MLPALFMSFLASFORKSLCVTAALV--GALAGVTLFSPVAILAGIVCGCLTALIOAFWG 230	QY	192 SFQRKSLCVTAALVGLAGVTLFSPVAILAGIVCGCLTALIOAFWG 240
DB	179 ALPALFLALVCQSGPMQGRMRPVLAALAVGGALAAVG--QTTLAILAGAAAGCL 229	DB	174 NLKGRH--VRAAVSGGVALA-FHLLNLTSVGIIAAALAGPLLSGWDG 218
RESULT 12		RESULT 13	
Y1755 ARCFU		Q8VM59 BRUAB	
ID	Y1755 ARCFU STANDARD; PRT; 219 AA.	ID	Q8VM59 BRUAB PRELIMINARY; PRT; 224 AA.
AC	O28519;	AC	Q8VM59; Q57B37;
DT	15-JUL-1998 (Rel. 36, Created)	DT	01-WAR-2002 (TRENBLrel. 20, Created)
DT	15-JUL-1998 (Rel. 36, Last sequence update)	DT	01-WAR-2002 (TRENBLrel. 20, Last sequence update)
DT	10-MAY-2005 (Rel. 47, Last annotation update)	DT	13-SEP-2005 (TRENBLrel. 31, Last annotation update)
DE	Hypothetical protein AF1755.	DE	Hypothetical protein (Azic family protein).
GN	OrderedLocusNames=AF1755;	GN	OrderedLocusNames=Brubi_1832;
OC	Archaeoglobus fulgidus.	OS	Brucella abortus.
OC	Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;	OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC	Archaeoglobaceae; Archaeoglobus.	OC	Brucellaceae; Brucella.
OX	NCBI_TaxID=2234;	OX	NCBI_TaxID=235;
NC	[1]	NC	[1]
RP	NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].	RP	NUCLEOTIDE SEQUENCE.
RC	STRAIN=VC-16 / DSM 4304 / ATCC 49558;	RC	STRAIN=544;
RC	MEDLINE=98049343; PubMed=9389475; DOI=10.1038/37052;	RC	MEDLINE=21952364; PubMed=11955619; DOI=10.1016/S0167-4781(01)00352-9;
RA	Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,	RA	Halling S.M., Zuerner R.L.;
RA	Ketchum K.A., Dodson R.J., Gwinn M.L., Hickey E.K., Peterson J.D.,	RA	"Evidence for lateral transfer to Brucellae: characterization of a
RA	Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,	RT	locus with a Tn-like element (tn2020)."
RA	Flieschmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.R.,	RL	Biochim. Biophys. Acta 1574:109-116(2002).
RA	Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.J.,	R	

RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 RC STRAIN=9-941 / Biovar 1;
 RX PubMed=15805518; DOI=10.1128/JB.187.8.2715-2726.2005;
 RA Halling S.M., Peterson-Burch B.D., Bricker B.J., Zuerner R.L.,
 RA Qing Z., Li L.-L., Kapur V., Alt D.P., Olsen S.C.;
 RT "Completion of the genome sequence of *Brucella abortus* and comparison
 RT to the highly similar genomes of *Brucella melitensis* and *Brucella*
 RT suis";
 RL J. Bacteriol. 187:2715-2726 (2005).
 DR EMBL; AF118548; AAL32283.1; -; Genomic DNA.
 DR EMBL; AB017223; AAX75147.1; -; Genomic_DNA.
 DR InterPro; IPR011606; AzlC_like.
 DR Pfam; PF03591; AzlC; 1.
 KW Complete proteome; Hypothetical protein.
 SQ SEQUENCE 224 AA; 23597 MW; 1576239416786FDC CRC64;

Query Match 24.7%; Score 308.5; DB 2; Length 224;
 Best Local Similarity 35.1%; Pred. No. 5.9e-17;
 Matches 78; Conservative 45; Mismatches 86; Indels 13; Gaps 6;
 Qy 15 FMEGCKDLPVIVISYIPVAFAGLNATRLGFSPLSFVFFSCIIYAGASQ-EVITAMLAAG 73
 Db 8 PFHGARSAPILIGLVPVAFAGTAAGSQGLALDSTALSMFSGANQAFFLSAVASGL 67
 Qy 74 SSLMIAALTVMAMDVHVLGYSRSLRIORLQSKTALMAFGLTDEVFAA---ATAKLV 130
 Db 68 PTIAIVAICAVA-SLRHILYGFVLRRLAGL--ASRLAFAGLGTDEVFAVTLNATEK-- 122
 Qy 131 RNNRRWSENWMIATFSSWSWVGTVIGAFSGSLGQYPAVEAALGFMPLPALFMSFLL 190
 Db 123 ---SKPDGWIIFGLAFPAWISWAATFFGAWMNILOAQFLQSLDALHFAFPALFLGLVW 179
 Qy 191 ASFOKQSLCVTAALVAGALAGTVLFSIPVAILAGIVCGCLTA 232
 Db 180 VTSARNVPMVAIAVIAVMPFLCL-NLPALAIPGAASAAALLA 220

RESULT 14
 Q8FYLS BRUSU
 ID Q8FYLS BRUSU PRELIMINARY; PRT; 224 AA.
 AC Q8FYLS;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE AzlC family protein.
 GN OrderedLocusNames=BR1853;
 OS Brucella suis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Brucellaceae; Brucella.
 OC NCBI_TaxID=29461;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=1330 / Biovar 1;
 RX MEDLINE=2224774; PubMed=12271122; DOI=10.1073/pnas.192319099;
 RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
 RA Read T.D., Dodson R.J., Nayana L.A., Brinkac L.M., Beanan M.J.,
 RA Daugherty S.C., DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R.,
 RA Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J.A.,
 RA Van Aken S.E., Riedmuller S., Tettelin H., Gill S.R., White O.,
 RA Salzberg S.L., Hoover D.L., Lindler L.E., Halling S.M., Boyle S.M.,
 RA Fraser C.M.;
 RT "The *Brucella suis* genome reveals fundamental similarities between
 RT animal and plant pathogens and symbionts";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153 (2002).
 DR EMBL; AE014291; AAN30748.1; -; Genomic_DNA.
 DR TIGR; BR1853; -;
 DR InterPro; IPR011606; AzlC_like.
 DR Pfam; PF03591; AzlC; 1.
 KW Complete proteome.
 SQ SEQUENCE 224 AA; 23597 MW; 1576239416786FDC CRC64;

Query Match 24.7%; Score 308.5; DB 2; Length 224;
 Best Local Similarity 35.1%; Pred. No. 5.9e-17;

Matches 78; Conservative 45; Mismatches 86; Indels 13; Gaps 6;
 Qy 15 FMEGCKDLPVIVISYIPVAFAGLNATRLGFSPLSFVFFSCIIYAGASQ-EVITAMLAAG 73
 Db 8 PFHGARSAPILIGLVPVAFAGTAAGSQGLALDSTALSMFSGANQAFFLSAVASGL 67
 Qy 74 SSLMIAALTVMAMDVHVLGYSRSLRIORLQSKTALMAFGLTDEVFAA---ATAKLV 130
 Db 68 PTIAIVAICAVA-SLRHILYGFVLRRLAGL--ASRLAFAGLGTDEVFAVTLNATEK-- 122
 Qy 131 RNNRRWSENWMIATFSSWSWVGTVIGAFSGSLGQYPAVEAALGFMPLPALFMSFLL 190
 Db 123 ---SKPDGWIIFGLAFPAWISWAATFFGAWMNILOAQFLQSLDALHFAFPALFLGLVW 179
 Qy 191 ASFOKQSLCVTAALVAGALAGTVLFSIPVAILAGIVCGCLTA 232
 Db 180 VTSARNVPMVAIAVIAVMPFLCL-NLPALAIPGAASAAALLA 220

RESULT 15
 Q6LZP2 METMP
 ID Q6LZP2 METMP PRELIMINARY; PRT; 238 AA.
 AC Q6LZP2;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE AzlC related protein.
 GN OrderedLocusNames=MWP0582;
 OS Methanococcus maripaludis.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 OC Methanococcaceae; Methanococcus.
 OC NCBI_TaxID=39152;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=S2 / LL;
 RX PubMed=15466049; DOI=10.1128/JB.186.20.6956-6969.2004;
 RA Hendrickson E.L., Kaul R., Zhou Y., Bovee D., Chapman P., Chung J.,
 RA Conway de Macario E., Dodsworth J.A., Gillett W., Graham D.E.,
 RA Hackett M., Haydock A.K., Kang A., Land M.L., Levy R., Lie T.J.,
 RA Major T.A., Moore B.C., Porat I., Palmeiri A., Rouse G.,
 RA Saenphimmachak C., Soell D., Van Dien S., Wang T., Whitman W.B.,
 RA Xia Q., Zhang Y., Larimer F.W., Olson M.V., Leigh J.A.;
 RT "Complete genome sequence of the genetically tractable
 RT hydrogeotrophic methanogen *Methanococcus maripaludis*.";
 RL J. Bacteriol. 186:6956-6969 (2004).
 DR EMBL; BX957220; CAF30138.1; -; Genomic_DNA.
 DR InterPro; IPR011606; AzlC_like.
 DR Pfam; PF03591; AzlC; 1.
 KW Complete proteome.
 SQ SEQUENCE 238 AA; 26219 MW; 70E25B7BBB74D035 CRC64;

Query Match 24.2%; Score 301.5; DB 2; Length 238;
 Best Local Similarity 32.6%; Pred. No. 2.3e-16;
 Matches 78; Conservative 46; Mismatches 96; Indels 19; Gaps 6;
 Qy 15 FMEGCKDLPVIVISYIPVAFAGLNATRLGFSPLSFVFFSCIIYAGASQ-EVITAMLAAGS 74
 Db 10 YFSGIKDAIPISIGLIPGAFVLAQKSMGIPVEISILMSLIIFAGASQFVGNVLIAGT 69
 Qy 75 SLMIAALTVMAMDVHVLGYSRSLRIORLQSKTALMAFGLTDEVFAAATAKLVNRNR 134
 Db 70 SSPEIVLTFTILNRLHFLMSSSLSQRIDYTKSKKLSLISFGVTDETFAVASLK---EEL 126
 Qy 135 RWSENWMIATFSSWSWVGTVIGAFSGSLGQYPAVEAALGFMPLPALFMSFLLASF 193
 Db 127 KUSPEFLGLNFTAFPAWNPFGTVLIGF---LAESIPKEIQSSMGISLYMFIQLLIPAV 182
 Qy 194 QRKQSL---CVTAALVAGALAGTVLFSIP-----VAILAGIVCGCLTALIQAFWOGAPDE 244
 Db 183 RRSKTKVLKFLVAVPFISSALTWPVPKFIETGTHIITIVASFIGA---KFIHGDGDE 238

Search completed: February 15, 2006, 12:23:33

Job time : 207.02 secs



GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 12:23:52 ; Search time 46.1096 Seconds
(without alignments)
439.292 Million cell updates/sec

Title: US-10-073-293A-4

Perfect score: 1248

Sequence: 1 MESPTPOPGSATFMGCK.....VCGCLTALIOAFWQAPDEL 245

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/1/iaa/5_COMB.pep.*

2: /cgn2_6/ptodata/1/iaa/6_COMB.pep.*

3: /cgn2_6/ptodata/1/iaa/H_COMB.pep.*

4: /cgn2_6/ptodata/1/iaa/PCFUS_COMB.pep.*

5: /cgn2_6/ptodata/1/iaa/RE_COMB.pep.*

6: /cgn2_6/ptodata/1/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1013.5	81.2	251	2	US-09-489-039A-10187
2	592	47.4	266	2	US-09-328-352-6525
3	246.5	19.8	247	2	US-09-489-039A-8478
4	229.5	18.4	230	2	US-09-710-279-268
5	225.5	18.1	242	2	US-09-134-001C-5154
6	214.5	17.2	255	2	US-09-107-532A-4920
7	209	16.7	345	2	US-09-252-991A-27947
8	199.5	16.0	263	2	US-09-328-352-4486
9	197	15.8	251	2	US-09-471-803A-3
10	197	15.8	251	2	US-10-608-504-3
11	197	15.8	251	2	US-09-605-703B-1840
12	194.5	15.6	206	2	US-09-134-000C-6789
13	187	15.0	224	2	US-09-583-110-4834
14	180	14.4	237	2	US-09-605-703B-2342
15	171	13.7	185	2	US-09-107-433-3341
16	146.5	11.7	230	2	US-09-543-681A-4173
17	143	11.5	288	2	US-09-489-039A-12218
18	122.5	9.8	247	2	US-09-328-352-6024
19	113.5	9.1	202	2	US-09-902-540-14110
20	102	8.2	306	2	US-09-328-352-5398
21	101.5	8.1	474	2	US-09-605-703B-2802
22	101	8.1	538	2	US-09-489-039A-9262
23	100	8.0	326	2	US-09-489-039A-11067
24	99.5	8.0	311	2	US-09-252-991A-19489
25	98.5	7.9	467	1	US-08-805-118-3
26	98.5	7.9	467	2	US-09-391-958-3
27	98	7.9	480	1	US-08-724-394A-9

28	97.5	7.8	441	2	US-09-950-071-2	Sequence 2, Appli
29	97	7.8	397	2	US-09-489-039A-12465	Sequence 12465, A
30	97	7.8	548	2	US-09-328-352-6605	Sequence 6605, Ap
31	96	7.7	362	2	US-09-902-540-12082	Sequence 12082, A
32	96	7.7	514	2	US-09-489-039A-11902	Sequence 11902, A
33	95	7.6	528	2	US-09-489-039A-9739	Sequence 9739, Ap
34	94.5	7.6	409	2	US-09-328-352-6044	Sequence 6044, Ap
35	94.5	7.6	549	2	US-09-115-150-4	Sequence 4, Appli
36	93.5	7.5	473	2	US-09-489-039A-9854	Sequence 9854, Ap
37	93	7.5	236	2	US-09-583-110-5253	Sequence 5253, Ap
38	93	7.5	473	2	US-09-303-518D-376	Sequence 376, App
39	92.5	7.4	235	2	US-09-583-110-5221	Sequence 5221, Ap
40	92.5	7.4	247	2	US-09-107-433-4191	Sequence 4191, Ap
41	92.5	7.4	411	2	US-09-489-039A-7912	Sequence 7912, Ap
42	92.5	7.4	510	2	US-09-605-703B-86	Sequence 86, Appl
43	92.5	7.4	561	2	US-09-252-991A-26010	Sequence 26010, A
44	92	7.4	327	2	US-09-252-991A-27652	Sequence 27652, A
45	91.5	7.3	426	2	US-09-583-110-4102	Sequence 4102, Ap

ALIGNMENTS

RESULT 1
US-09-489-039A-10187
; Sequence 10187, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 10187
; LENGTH: 251
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-10187

Query Match	81.2%	Score	1013.5;	DB 2;	Length	251;
Best Local Similarity	81.0%	Pred. No.	2.8e-104;			
Matches	200;	Conservative	16;	Mismatches	26;	Indels 3; Gaps 1;
Qy	1	MESPTPOPA---	PGSATFMGCKD	SLPIVISYIPVAF	PGFNATRLG	FSPLSFVFCII 57
Db	4	MENPAPLT	CALPERVATV	GVGKDSLPIVISY	LPVAFPGFNATRL	GFTPLSFVFCII 63
Qy	58	YAGASQPVIT	AMLAAGSSLI	WIAALTVMMDVR	HVLYGPSLSR	RIORLQKSKTALWAFGL 117
Db	64	YAGASQPVIT	AMLAAGSSLI	WIAALTVMMDVR	HVLYGPSLSR	RIORLQKSKTALWAFGL 123
Qy	118	TDEVF	AAATAKLV	YRNRRN	SENW	MGIAFSWSWVFGT
Db	124	TDEVF	AAATAKLV	YRNRRN	SENW	MGIAFSWSWVFGT
Qy	178	GFMLPAL	FMSFL	LAASFOR	KOSLCV	TALVAGLV
Db	184	GFMLPAL	FMSFL	LAASFOR	KOSLCV	TALVAGLV
Qy	238	WQAGP	DE 244			
Db	244	LKGM	PDE 250			
RESULT 2						
US-09-328-352-6525						
; Sequence 6525, Application US/09328352						
; Patent No. 6562958						
; GENERAL INFORMATION:						

```

; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6525
; LENGTH: 266
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6525

```

Query Match 47.4%; Score 592; DB 2; Length 266;
Best Local Similarity 48.1%; Pred. No. 2e-57;
Matches 111; Conservative 53; Mismatches 67; Indels 0; Gaps 0;

Qy	7	QAPGSA	TFME	GCKDS	LP	IV	LSY	YP	VA	FA	FG	LN	TR	LG	FS	PL	ES	VP	FS	CI	YAG	AS	Q	P	V	I	66	
Db	26	QMAA	QTAT	FW	GAKD	SO	AI	VL	TV	LP	VF	AF	GS	Q	Q	FG	T	P	WE	A	FF	LS	CM	YAG	AS	Q	FLV	85
Qy	67	TAM	LA	AG	SS	L	W	I	A	L	T	W	M	D	V	R	H	L	Y	G	P	S	L	R	S	R	I	126
Db	86	VALL	AG	SS	I	L	T	W	I	A	L	T	W	D	V	R	H	L	Y	G	P	S	L	R	S	R	I	145
Qy	127	AK	L	V	R	N	R	N	R	S	E	N	W	I	G	A	F	S	W	S	W	S	V	G	T	V	I	186
Db	146	IQL	S	Q	R	Q	W	S	E	N	W	I	G	A	F	S	W	S	W	S	L	G	L	G	L	F	A	205
Qy	187	SFL	L	A	S	F	Q	R	K	O	S	L	C	V	T	A	L	V	A	L	G	A	G	V	T	L	F	237
Db	206	SFL	L	A	S	F	Q	R	K	S	L	V	A	L	V	A	L	G	A	G	V	T	L	F	S	I	A	256

RESULT 3
US-09-489-039A-8478
; Sequence 8478, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489.039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 8478
; LENGTH: 247
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
; US-09-489-039A-8478

Query Match	19.8%;	Score	246.5;	DB	2;	Length	247;
Best Local Similarity	28.6%;	Pred. No.	4.5e-19;				
Matches	67;	Conservative	46;	Mismatches	106;	Indels	15;
Gaps	6;						
Qy	18	GCKDSLPTIVISYIPVAPAFGLNATLGLSPLESVFSCIIYAGASQFVITAMLAGSSLW	77				
Db	10	GASAIMPLCTGDPFSPFIVGALSVSAGMSVWQSTAWSAIVISAGSOMLANMLKGTATLG	69				
Qy	78	IAALTVMAMDVHRVLYGSPLSRRIQRLQKSKTALWAFGLTDEVPAAATAKLVNRNRWS	137				
	:						
Db	70	VIIPTTLILNLRHVLYSASI-SGTVREASPFKCKCFMSYALTDEVY-ATTVMEMEGNKKK	127				
Qy	138	ENNMIGTAFSSWSNVFCTVIGASGSGLLQGYPAVEA-ALGEWLPALPMSFLLASFOK	196				
Db	128	YLFYGSAMITFWALWLDADFALVGVA---SFPHEIKYGDGFAMVAAPFAIVVP--QIK	181				
Qy	197	QSLCVTAALVGALAGTVLFSIP-----VAILAGIVCGCLTALTQAFWQGPAPDE	244				
Db	182	SOACTVAAVVAASGVLLVLPYSYLGIVVAASVVLGAGLCVYDLAERKOMAKTE	235				

```

RESULT 4
US-09-710-279-268
; Sequence 268, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 268
; LENGTH: 230
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-09-710-279-268

```

	Query Match	18.4%;	Score 229.5;	DB 2;	Length 230;
	Best local Similarity	28.6%;	Pred. No. 3.2e-17;		
	Matches 65;	Conservative 47;	Mismatches 102;	Indels 13;	Gaps 6;
Qy	14	TFMEGCKDSLPIVTSYIPVAFAGFNATRLGFPSPLESVFFSCIIYAGASQPVITAMLAAG	73		
Db	5	TFKGVKECTPTLLGYACVGLSGFIVASQNFVLEILLCLIIYAGAQFIICTLVIAG	64		
Qy	74	SSLWIAALTVMAMDVRHVLGYPSLRSRIIQRLQKSKTALWAFGLTDEVFAAATAKLVRNN	133		
Db	65	TPISAIVLITILVNSRMFLLSMTLAPKNYKQGFNWRVGLGTL-LTDETFGVAITFVVK-G	122		
Qy	134	RWSENWMIQIAPSSWSGSSVPGTVIGAFSGSGLLQGYPAVEAALG--FMLPALFMSFLLA	191		
Db	123	EKINDRWLHGLNITAYLFTWVSCVIGAFGE-----YISPDALGLDFAITAMFIFLCIS	177		
Qy	192	SFQ--RQKSLCVTAALGALAGVTLF--SIPVAILAGIVCGCLTALI	234		
Db	178	OFEGIGKESRIRIYTVITVCVIVMMMLLSLILPSYLAIIIAIVALL	224		

```

RESULT 5
US-09-134-001C-5154
; Sequence 5154, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GPC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5154
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-5154

```

Query Match	18.1%;	Score 225.5;	DB 2;	Length 242;
Best Local Similarity	28.6%;	Pred. No. 9.5e-17;		
Matches	66;	Conservative	39;	Mismatches 89;
			Indels	37;
			Gaps	7;
Qv	14	TFMEGCKDSLPIVTSYIPVAFGLNATRLGSPLESVFSCIIYAGASQFVTAMLAAG	73	

[illegible]

```
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4486
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4486

Query Match      16.0%; Score 199.5; DB 2; Length 263;
Best Local Similarity 29.0%; Pred. No. 8.3e-14;
Matches 65; Conservative 41; Mismatches 101; Indels 17; Gaps 6;

Qy 15 FMEGCKDSLPIVISIPVAFAGLNATRLGSPLESVFFSCIIYAGASQFVITAMLAAGS 74
Db 45 FLRGAILDILPISIVIPWAILAGSMAVHAGLSFYKALAMSGIVPAGAAQLVLSWMEGA 104
Qy 75 SLWIAALTVMAMDVRHVLVYGPSLRSRIQRLOKSKTALMAFGLTDEVPAAATAKLVRNNR 134
Db 105 SLLTIYVTFEFLTAQHFIYALTLRND-ISLPLSKRLTLGLFLLTDELFAVSV-----NE 158
Qy 135 RWSNWMIGIAFSSWSWVFTVIGAFSGSL--LQGYPAVEAALGFMLPALFMSFLLAS 192
Db 159 KKHPOYLFAGLCFLFYFWVFSLVGILLATLPNLLNY-----HLDFSIIAIFVAMIVPM 213
Qy 193 FORQSLCVTAALVAGALAGVTL--FSIPVAILAGIVCGCLTALI 234
Db 214 CKGRP--VMAGILMTCVSGFVLKPFHIEGALLISGLLGMFTAVI 255

RESULT 9
US-09-471-803A-3
; Sequence 3, Application US/09471803A
; Patent No. 6613545
; GENERAL INFORMATION:
; APPLICANT: KENNERKNECHT, NICOLE
; APPLICANT: SAHM, HERMANN
; APPLICANT: EGGELE, LOTHAR
; APPLICANT: PFEFFERLE, WALTER
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE CODING FOR THE EXPORT OF
; TITLE OF INVENTION: BRANCHED CHAIN AMINO ACIDS, PROCESS FOR THE
; TITLE OF INVENTION: ISOLATION THEREOF AND USE THEREOF
; FILE REFERENCE: 21123/265496/MAS
; CURRENT APPLICATION NUMBER: US/09/471,803A
; CURRENT FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: DE 199 51 708.8
; PRIOR FILING DATE: 1999-10-27
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 251
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
; FEATURE:
; OTHER INFORMATION: ATCC14752
US-09-471-803A-3

Query Match      15.8%; Score 197; DB 2; Length 251;
Best Local Similarity 27.8%; Pred. No. 1.5e-13;
Matches 65; Conservative 47; Mismatches 92; Indels 30; Gaps 9;

Qy 17 EGCKDSLPIVISIPVAFAGLNATRLGSPLESVFFSCIIYAGASQFVITAMLAAGSSL 76
Db 34 QGLKTSLAAGLGMYPGIGAFGLLVIOGYEWWAAPFLSGLIFAGSTEMLVAVVGAAPL 93
Qy 77 WIAALTVMAMDVRHVLVYGPSLRSRIQRLOKSKTALM--AFGLTDEVPAAATAKLVRNNR 134
Db 94 GAIALTTLLVNFVRHVFYAFSPPLHV-----KNPIARFYSVFALIDEAYAVTAAR----PA 145
Qy 135 RWSNWMIGIAFSSWSWVFTVIGAFSGSLGQYPAVEAALGFMLPALFMSFLLASFO 194
Db 146 GWSAWRLISMQIAFHSYWF-----GGLTGVAIAELIPFEIKGLEFALCSLFVTITLDSR 201
Qy 195 RKQSLCVTAALVAGALAGVTLFSIPVAILAG-----IVCGCLTALIQAFWQ 240
Db 202 TKKQI---PSLL--LAGLS-FTIALVVPQALFAALLIFGLLT--IRYFFLG 247

RESULT 11
US-09-605-703B-1840
; Sequence 1840, Application US/09605703B
; Patent No. 6962989
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberhauer, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING NOVEL
; TITLE OF INVENTION: PROTEINS
; FILE REFERENCE: BGI-129CP
; CURRENT APPLICATION NUMBER: US/09/605,703B
; CURRENT FILING DATE: 2000-06-27
; PRIOR APPLICATION NUMBER: 60/142,764
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: 60/152,318
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 2934
```

; SEQ ID NO 1840
; LENGTH: 251
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-605-703B-1840

Query Match 15.8%; Score 197; DB 2; Length 251;
Best Local Similarity 27.8%; Pred. No. 1.5e-13;
Matches 65; Conservative 47; Mismatches 92; Indels 30; Gaps 9;

Qy 17 EGKDSLPIVISYIPVAFAGLNATRLGSPLESVFSCIIYAGASQFVTITMLAAGSSL 76
Db 34 QGLKTSLAAGLGMTPIGIAFGLLVIOGYEWAAPLFSLGIFAGSTEMLVIALVGAAPL 93
Qy 77 WIAALTVMAMDVHVLVGPSSRSRIORLQSKTALW--AFGLTDEVFVAAATKLVNRNR 134
Db 94 GAIALTLLVNFVRHVFVAFSPPLHV---KNPIARFVSVPALIDEAYAVTAAR----PA 145
Qy 135 RWSNNWIGIAFSSWSNVFVTIGAFSGGLQGYPAVEAALGFMLPALFMSFLASFO 194
Db 146 GWSAWRLISMQIAFHSYVVF---GGLTGVAIAELIPPEIKGLEFALCSLFTVLTLSR 201
Qy 195 RKQSLCVTAALVAGLAGVTLFSPVAILAG-----IVCGCLTALIQAFWQ 240
Db 202 TKQI---PSLL--LAGLS-FTIALVVPQALFAALLIFLGLT--IRYFFLG 247

RESULT 12
US-09-134-000C-6789
; Sequence 6789, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6789
; LENGTH: 206
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-6789

Query Match 15.6%; Score 194.5; DB 2; Length 206;
Best Local Similarity 31.1%; Pred. No. 2.1e-13;
Matches 65; Conservative 29; Mismatches 92; Indels 23; Gaps 7;

Qy 44 GSPLESVFSCIIYAGASQFVTITMLAAGSSLWIAALTVMAMDVHVLVGPSSRSRI-I 102
Db 7 GSPSLIVSAMSFPFIAGSAQFVTIVSMLTGGSPILSVLATFLVNARMILGMTIAPYFKA 66
Qy 103 QRLQSKTALWAFQ--LTDSVEFAATATKLVNRNRWSENWMIAGFSSWSNVFGTVIGA 160
Db 67 ESIGKN---LW-LGTLTDSFALGWMKLNHTKRLSFEWNAANLISYATWVFSTIIA 122
Qy 161 FSGSGLLQGYPAVEAALG--FMLPALFMSFLASQFQKSLCVTAALVAGLAGVTLFSP 218
Db 123 YLGR-----FIANPQALGLEFAVAVAMPFGLLYLQIISDRSMKIALQJLWLIFFGLMSIG 177
Qy 219 -----VAIAGIVCGCLTALIQAFW 238
Db 178 LIFIPSNLIVLVVTLLIGGIRVMIKHAFF 206

RESULT 13
US-09-583-110-4834
; Sequence 4834, Application US/09583110
; Patent No. 6699703

; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics
; FILE REFERENCE: PATH00-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 4834
; LENGTH: 224
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; NAME/KEY: UNSURE
; LOCATION: (188)
; OTHER INFORMATION: Identity of amino acid sequences at the above locations are unknown
US-09-583-110-4834

Query Match 15.0%; Score 187; DB 2; Length 224;
Best Local Similarity 25.6%; Pred. No. 1.6e-12;
Matches 60; Conservative 48; Mismatches 96; Indels 30; Gaps 10;

Qy 17 EGKDSLPIVISYIPVAFAGLNATRLG---FSPLESVFSCIIYAGASQFVTITMLAAG 73
Db 1 EGAQAAMPTALGYVSGIACG---TIGAPYVTPVEMGLMSLFYAGSAQFAMLLIIVVQ 56
Qy 74 SSWIAALTVMAMDVHVLVGPSSRSRIORLQSKTALW-----AFGLTDEVFVAAATK 128
Db 57 APVAATAMTVFLINLR--LFLLSLHASTYR----HTSLWNMGMSILLIDETVGLMGE 110
Qy 129 LVNRNRWSENWMIAGFSSWSNVFVTIGAFSGGLQGYPAVEAALGFMLPALFMSF 188
Db 111 LAHTD-KVNDPMWMMHGNLSYVAVFVGTVGTALG-GLLPN-PEI-FGLDFALVGMFIGI 166
Qy 189 LLASFQKQS-----LCVTAALVAGLAGVTLFSPVAILAGIVCGCLTALI 234
Db 167 FASQFQMMQRRIPVRNLLIILXVAVSVFLLTLTVSQSLAVLFAMLLIGSCMGVV 220

RESULT 14
US-09-605-703B-2342
; Sequence 2342, Application US/09605703B
; Patent No. 6962989
; GENERAL INFORMATION:
; APPLICANT: Pompeius, Markus
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberkauer, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING NOVEL
; FILE REFERENCE: BGI-129CP
; CURRENT APPLICATION NUMBER: US/09/605,703B
; CURRENT FILING DATE: 2000-06-27
; PRIOR APPLICATION NUMBER: 60/142,764
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: 60/152,318
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 2934
; SEQ ID NO 2342
; LENGTH: 237
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-605-703B-2342

Query Match 14.4%; Score 180; DB 2; Length 237;
Best Local Similarity 27.8%; Pred. No. 1.1e-11;

Matches 59; Conservative 39; Mismatches 88; Indels 26; Gaps 6;

QY 18 GKCDSLPIVISYIPVAFAGNATRLGFSPLSFVFFSIIYAGASOFVITAMLAAGSSLW 77
Db 12 GIGETLTVGLGLPLGLAFGLLWQVGFAMWTFPIFSFLVIYAGSMFLAIGMTAGIGP 71
QY 78 IAAITWAMDVRHVLVGPSSL-RSRIIQRLOKSKT--ALMAFGLTDEVFAAATAKLVRNRR 134
Db 72 SAAVAGFMVNFRIIFYGLTTPRHRI-----KSGAGRAYSTYALTDESYAIVSA---RPPG 123
QY 135 RWSNWMIGIAGFSNWSWFGTGVIGAFSGSLGQYPAVEAALGFMLPALEMSFLLASFQ 194
Db 124 DISGTRVLTVQIICQALWVTPGIIGALVG---QVLPLDLKGMDFALTALFVVLAWAEAFK 179
QY 195 RKQSLCVTRALVAGALAGVTLFSIPVAILAGIV 226
Db 180 NNDKY-----SLPLFAVVVLVSGFV 200

RESULT 15
US-09-107-433-3341
; Sequence 3341, Application US/09107433
; Patent No. 6800744
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3341:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 185 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...185
; SEQUENCE DESCRIPTION: SEQ ID NO: 3341:
US-09-107-433-3341

Query Match 13.7%; Score 171; DB 2; Length 185;
Best Local Similarity 28.0%; Pred. No. 7.4e-11;


```

Db 181 LPALFMSLLASFORKQSLCVTAALVGLAGVTLFSPVAILAGVCGCLTALIQAFWQ 240
;
Qy 241 APDEL 245
;
Db 241 APDEL 245
;

RESULT 2
US-10-724-972A-6721
; Sequence 6721, Application US/10724972A
; Publication No. US20040147734A1
; GENERAL INFORMATION:
; APPLICANT: Doucette-Stamm, Lynn
; APPLICANT: Bush, David
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/10/724,972A
; PRIOR FILING DATE: 2003-12-01
; PRIOR FILING DATE: 1999-11-29
; PRIOR FILING DATE: 1999-11-29
; PRIOR FILING DATE: 1998-08-13
; PRIOR FILING DATE: 1997-11-08
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 7544
; SEQ ID NO 6721
; LENGTH: 242
; TYPE: PRT
; ORGANISM: S.epidermidis
US-10-724-972A-6721

Query Match 18.1%; Score 225.5; DB 4; Length 242;
Best Local Similarity 28.6%; Pred. No. 3.2e-14;
Matches 66; Conservative 39; Mismatches 89; Indels 37; Gaps 7;

Qy 14 TFMGCKDSLPIVISYIPVAFAGLNATRLGFSPLSVFFSCIIYAGASQFVITAMLAAG 73
;
Db 17 TFKQVKECIPITLLGYAGVGLSFGIVAASQNFVLEIILLCLIIYAGAAQFIITLVIA 76
;
Qy 74 SSLIAALTVMAMDVRHVLYGSLRSRIIQRLQSKTALWAFGLTDEVPAAATAKLVN 133
;
Db 77 TPISAIVLTLLIVNSRMFLSMTLAPNYKQYGFWRVGLGLT-LTDETFGVAITPYVK-G 134
;
Qy 134 RWSENMMIGTAFSSWSWVFTGIVGAFSGSGLLQGYPAVEAALG--FMLPALFMSFLA 191
;
Db 135 EKINDRWLHUNITAYLFTWVSCVIGAFGE-----YISNPDALGLDPAITAMFIFLCIS 189
;
Qy 192 SFQ--RKQSL-----CV-----TAALVGLAGVTL 214
;
Db 190 QFEGIKKSRLAIYIIVCIVVMMLLSSILPSYVAIIIAIVAALLGVVM 240
;

RESULT 3
US-09-738-626-3789
; Sequence 3789, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125

```

```

; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patentin ver. 3.0
; SEQ ID NO 3789
; LENGTH: 251
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-3789

Query Match 15.8%; Score 197; DB 3; Length 251;
Best Local Similarity 27.8%; Pred. No. 2.5e-11;
Matches 65; Conservative 47; Mismatches 92; Indels 30; Gaps 9;

Qy 17 EGCKDSLPIVISYIPVAFAGLNATRLGFSPLSVFFSCIIYAGASQFVITAMLAAGSSL 76
;
Db 34 QGLKTSLAAGLGMYPGIGIAFGLLVIIQYGYEWAAAPLFSGLIFAGSTEMLVIALVVGAPL 93
;
Qy 77 WIAALTVMAMDVRHVLYGSLRSRIIQRLQSKTALW--AFGLTDEVFAAATAKLVNRR 134
;
Db 94 GAIALTTLLVNFHRVYFAFSFPLHV-----KNPIARFYSVPALIDEAYAVTAAR----PA 145
;
Qy 135 RWSENMMIGTAFSSWSWVFTGIVGAFSGSGLLQGYPAVEAALGFMPLPMSFLLASFQ 194
;
Db 146 GWSAURLISMQIAFHSYVWF---GGLTGVAIAELIPEIKGLEFALCFLPVTILTDSCR 201
;
Qy 195 RKQSLCVTAALVGLAGVTLFSPVAILAG-----IVCGCLTALIQAFWQ 240
;
Db 202 TKKQI---PSLL--LAGLS-FTIALVVPQALFAALLIFLGLLT--IRYFFLG 247
;

RESULT 4
US-10-608-504-3
; Sequence 3, Application US/10608504
; Publication No. US20040014123A1
; GENERAL INFORMATION:
; APPLICANT: KENNERKNECHT, NICOLE
; APPLICANT: SAHM, HERMANN
; APPLICANT: EGGLING, LOTHAR
; APPLICANT: PFEFFERLE, WALTER
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE CODING FOR THE EXPORT OF
; TITLE OF INVENTION: BRANCHED CHAIN AMINO ACIDS, PROCESS FOR THE
; TITLE OF INVENTION: ISOLATION THEREOF AND USE THEREOF
; FILE REFERENCE: 7601/80525
; CURRENT APPLICATION NUMBER: US/10/608,504
; CURRENT FILING DATE: 2003-06-30
; PRIOR APPLICATION NUMBER: US 09/471,803
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: DE 199 51 708.8
; PRIOR FILING DATE: 1999-10-27
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 3
; LENGTH: 251
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
; FEATURE:
; OTHER INFORMATION: ATCC14752
US-10-608-504-3

Query Match 15.8%; Score 197; DB 4; Length 251;
Best Local Similarity 27.8%; Pred. No. 2.5e-11;
Matches 65; Conservative 47; Mismatches 92; Indels 30; Gaps 9;

Qy 17 EGCKDSLPIVISYIPVAFAGLNATRLGFSPLSVFFSCIIYAGASQFVITAMLAAGSSL 76
;
Db 34 QGLKTSLAAGLGMYPGIGIAFGLLVIIQYGYEWAAAPLFSGLIFAGSTEMLVIALVVGAPL 93
;

```

Qy 77 WIAALTVMAMDVHVLGSPSLRSRIORLQSKTALM--AFGLTDEVFAAATAKLVRNR 134
 Db 94 GAIALTILLVNRFRVFAFSPPLHV-----KNPIARFVSFALIDEAYAVTAAR----PA 145
 Qy 135 RWSNWMGIAFSSWSWVFTVIGAFSGSLGLOGYPAVEAALGFMLPALFMSFLLASFQ 194
 Db 146 GWSAWRLISMGIATFHSYWF-----GGLTGVAIAELIPPEIKGLEFALCSLFVTLTLDSCR 201
 Qy 195 RKQSLCVTAALVAGLAGVTLPFIPVAILAG-----IVCGCLTALIOAFWOG 240
 Db 202 TKQOI---PSLL--LAGLS-FTIALVVPFGQALFAALLIFGLT--IRYFFLG 247

RESULT 5

US-10-335-977-6207
 ; Sequence 6207, Application US/10335977
 ; Publication No. US20040052799A1

GENERAL INFORMATION:

APPLICANT: DOUGLAS SMITH et al
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
 RELATING TO HELICOBACTER PYLORI FOR
 DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031

CORRESPONDENCE ADDRESS:

ADDRESSER: LAHIVE & COCKFIELD
 STREET: 28 State Street
 CITY: Boston
 STATE: Massachusetts
 COUNTRY: USA
 ZIP: 02109-1875

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660
 COMPUTER: IBM PC Compatible
 OPERATING SYSTEM: Windows NT 4.0
 SOFTWARE: UNIX

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/335,977
 FILING DATE: 30-Dec-2002

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/993,002
 FILING DATE: 17-DEC-1997

ATTORNEY/AGENT INFORMATION:

NAME: Mandragouras, Amy E.
 REGISTRATION NUMBER: 36,207
 REFERENCE/DOCKET NUMBER: GTN-018
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617)227-7400
 TELEFAX: (617)742-4214

INFORMATION FOR SEQ ID NO: 6207:

SEQUENCE CHARACTERISTICS:
 LENGTH: 229 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: YES
 ORIGINAL SOURCE:
 ORGANISM: Helicobacter pylori
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (B) LOCATION 1...229

SEQUENCE DESCRIPTION: SEQ ID NO: 6207:

US-10-335-977-6207

Query Match 15.7%; Score 195.5; DB 4; Length 229;
 Best Local Similarity 27.0%; Pred. No. 3.2e-11;
 Matches 62; Conservative 42; Mismatches 85; Indels 41; Gaps 9;

Qy 15 FMEGCKDSLPIVSIYIPVAFAGLNATRLGFSPLSFVFCIIYAGASQFVITAML 70
 Db 5 FLAKFKADPHTTISIFIGYLLMGMTFGMLLAQQGYDKVALFMSLFIYAGAIQFVAITLL 64
 Qy 71 AAGSSLIAALTVMAMDVHVLGSPSLRSRIORLQSKTALM-----AFGLTDEVFAA- 124

Db 65 SAQASLMNVVIVSLVNNARQTCYALSM-----LDRFKNTKWLRYLAHALTDETFA 117
 Qy 125 ---ATAKLVNRNRWSNMWIGIAFSSWSWVFTVIGAFSGSL---LQGYPAVEAALG 178
 Db 118 NLTPAKKGVNET-----DFMFSILLNHSWIFGSLVSGSLVSHFSFDTQ3-----ME 165
 Qy 179 FMLPALFMSFLLASFORKQS-----LCVTAALVG-ALAGVTLPFIPVAIL 222
 Db 166 FVMTAIFVLFMQYKENTNHNKAWLGIAIAVAVCLALFGTEYFLIALVL 215

RESULT 6

US-10-501-282-2600
 ; Sequence 2600, Application US/10501282
 ; Publication No. US20050203280A1

GENERAL INFORMATION:

APPLICANT: MCMICHAEL, JOHN CALHOUN
 APPLICANT: ZAGURSKY, ROBERT JOHN
 APPLICANT: RUSSELL, DAVID PARRISH
 APPLICANT: FLETCHER, LEAH DIANE

TITLE OF INVENTION: ALLOIOCCOCUS OTITIDIS OPEN READING FRAMES (ORFS) ENCODING
 FILE OF INVENTION: POLYPEPTIDE ANTIGENS, IMMUNOGENIC COMPOSITIONS AND USES THEREOF
 FILE REFERENCE: AM100780 L2
 CURRENT APPLICATION NUMBER: US/10/501,282

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/333,777
 FILING DATE: 2004-07-09

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/426,742
 FILING DATE: 2002-11-18

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US02/36123
 FILING DATE: 2002-11-25

NUMBER OF SEQ ID NOS: 6653

SOFTWARE: PatentIn version 3.2

SEQ ID NO 2600

LENGTH: 226

TYPE: PRT

ORGANISM: Alloiooccus otitidis

US-10-501-282-2600

Query Match 15.1%; Score 189; DB 5; Length 226;
 Best Local Similarity 23.9%; Pred. No. 1.4e-10;
 Matches 55; Conservative 52; Mismatches 99; Indels 24; Gaps 6;

Qy 15 FMEGCKDSLPIVSIYIPVAFAGLNATRLGFSPLSFVFCIIYAGASQFVITAML 74
 Db 2 FKKVLKPAFPVMVSYLVLSVCGIVSFQVGTPLQILLTSVAVLYSGSGQFLLAGLYAGA 61
 Qy 75 SLWIAALTVMAMDVHVLGSPSLRSRIORLQSKTALMVF-----GLTDEVFAAATAKL 130
 Db 62 SLVSIITLTAFLGLRFLVMSMSSSRHV-----RQKTTWDFPFSSMTISDESFGVNTWFS 116
 Qy 131 RNNRWSNMWIGIAFSSWSWVFTVIGAFSGSLGLOGYPAVEAALGFMLPALFMSFLL 190
 Db 117 QPD--WTADHALALNLLNYGIWLGSLGAL-----LVSVVDLDTSIISYGLTAMFCMTV 170
 Qy 191 ASFORKQSLCVTAALVAGLAGVTLPFIPVAIL---AGIVCGCLTALIOAF 237
 Db 171 EQFVDY-----LYAGLISVVVFTIALVILQNSLGIIVGALLASLIGF 214

RESULT 7

US-10-156-761-10733
 ; Sequence 10733, Application US/10156761
 ; Publication No. US20030119018A1

GENERAL INFORMATION:

APPLICANT: OMURA, SATOSHI
 APPLICANT: IKEDA, HARUO
 APPLICANT: ISHIKAWA, JUN
 APPLICANT: HORIKAWA, HIROSHI
 APPLICANT: SHIBA, TADAYOSHI
 APPLICANT: SAKAKI, YOSHIYUKI
 APPLICANT: HATTORI, MASAHIRO

; PRIOR FILING DATE: 2001-03-27
 ; NUMBER OF SEQ ID NOS: 4979
 ; SOFTWARE: SeqMin99, version 1.03
 ; SEQ ID NO 96
 ; LENGTH: 218
 ; TYPE: PRT
 ; ORGANISM: Streptococcus pneumoniae
 ; FEATURE:
 ; OTHER INFORMATION: conserved hypothetical protein
 ; OTHER INFORMATION: Cellular location: membrane
 ; OTHER INFORMATION: Similar to strain R6 sequence 15902189 (2.E-55)
 US-10-472-928-96

 Query Match 14.3%; Score 178; DB 5; Length 218;
 Best Local Similarity 26.3%; Pred. No. 1.8e-09;
 Matches 60; Conservative 47; Mismatches 91; Indels 30; Gaps 11;

 QY 23 LPVISIPVAPAFGLNATRLG---FSPLESVPFSCIYAGASQFVTMTMLAAGSSLIWA 79
 DB 1 MPTALGVYVIGLACGI---IGAPYVTPVENGLMSLFYAGSAQFAMALITVQAPVAAI 56

 QY 80 ALTVMAMDMVRHLYGPSLRRIIQRLOKSKTALW----AFGLTDEVFAAATAKLVRNNR 134
 DB 57 AMTVFLNLR--LFELSLHASTYFR---HTSLWNMGMSILTDITYGVLMGELAHTD- 109

 QY 135 RWSNMMIGIAPSSWSWVFTVIGAFSGGLQGYPAVEAALGFMLPALFMSFLLASFO 194
 DB 110 KVPMMWGHNNLSYVAFVGTGVTALG-GLLPN-PEI-FGLDFALVGMFIFGIFASQFQ 166

 QY 195 RKQ-----SLCVTAALVGA--LAGVTLSFIPVAILAGIVCGCLTAII 234
 DB 167 IMORRIPVRNLLIILAVAVSFLLITVMSQSLAVLPFATLIGCSMGVV 214

 RESULT 11
 US-10-501-282-2598
 ; Sequence 2598, Application US/10501282
 ; Publication No. US20050203280A1
 ; GENERAL INFORMATION:
 ; APPLICANT: MICHAEL, JOHN CALHOUN
 ; APPLICANT: ZAGURSKY, ROBERT JOHN
 ; APPLICANT: RUSSELL, DAVID PARRISH
 ; APPLICANT: FLETCHER, LEAH DIANE
 ; TITLE OF INVENTION: ALLOIOCCOCUS OTITIDIS OPEN READING FRAMES (ORFS) ENCODING
 ; TITLE OF INVENTION: POLYPEPTIDE ANTIGENS, IMMUNOGENIC COMPOSITIONS AND USES THEREOF
 ; FILE REFERENCE: AM100780 L2
 ; CURRENT APPLICATION NUMBER: US/10/501,282
 ; CURRENT FILING DATE: 2004-07-09
 ; PRIOR APPLICATION NUMBER: 60/333,777
 ; PRIOR FILING DATE: 2001-11-29
 ; PRIOR APPLICATION NUMBER: 60/426,742
 ; PRIOR FILING DATE: 2002-11-18
 ; PRIOR APPLICATION NUMBER: PCT/US02/36123
 ; PRIOR FILING DATE: 2002-11-25
 ; NUMBER OF SEQ ID NOS: 6653
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 2598
 ; LENGTH: 214
 ; TYPE: PRT
 ; ORGANISM: Alloiooccus otitidis
 US-10-501-282-2598

 Query Match 13.8%; Score 172; DB 5; Length 214;
 Best Local Similarity 23.7%; Pred. No. 6.9e-09;
 Matches 53; Conservative 43; Mismatches 94; Indels 24; Gaps 6;

 QY 26 VISIPVAPAFGLNATRLGSPLESVPFSCIYAGASQFVTMTMLAAGSSLIWAALTVA 85
 DB 1 MVSVLVLSVCGIVSFQVGTFTQLITLSAVLYSGSGQFLLAGLYGAGASLSVSIITLAP 60

 QY 86 MDRHLYGPSLRRIIQRLOKSKTALWAP----GLTDEVFAAATAKLVRNNRWSNMM 141
 DB 61 LGLRFLVMSSSSRHV-----RQKTTWDFPFSSMTISDESFGVNTVMFSQPD--WTADHA 113

QY 142 IGIAFSSWSWVFTVIGAFSGGLQGYPAVEAALGFMLPALFMSFLLASFOKQSLCV 201
 DB 114 LALNLLNYGIWVLGSGAL---LVSVDLDTISIYGLTAMFICTMTVSGFVDRY---- 165

 QY 202 TAAIVGALAGVTLSFIPVAIL---AGIVCGCLTALIOAP 237
 DB 166 --YLYAGLSVVFETIILVILQNSLGIWVGLLASLIGF 202

 RESULT 12
 US-10-617-320-3341
 ; Sequence 3341, Application US/10617320
 ; Publication No. US20050136404A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn A Doucette-Stamm and David Bush
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
 ; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
 ; THERAPEUTICS
 ; NUMBER OF SEQUENCES: 5206
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
 ; STREET: 100 Beaver Street
 ; CITY: Waltham
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02354
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: CD/ROM ISO9660
 ; COMPUTER: <Unknown>
 ; OPERATING SYSTEM: <Unknown>
 ; SOFTWARE: <Unknown>
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/617,320
 ; FILING DATE: 10-Jul-2003
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/107,433
 ; FILING DATE: 30-Jun-1998
 ; APPLICATION NUMBER: 60/ 085131
 ; FILING DATE: May 12, 1998
 ; APPLICATION NUMBER: 60/051553
 ; FILING DATE: July 2, 1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Ariniello, Pamela Deneke
 ; REGISTRATION NUMBER: 40,489
 ; REFERENCE/DOCKET NUMBER: GTC-011
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (781)893-5007
 ; TELEFAX: (781)893-8277
 ; INFORMATION FOR SEQ ID NO: 3341:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 185 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; HYPOTHETICAL: YES
 ; ORIGINAL SOURCE:
 ; ORGANISM: Streptococcus pneumoniae
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (B) LOCATION 1...185
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 3341:
 US-10-617-320-3341

 Query Match 13.7%; Score 171; DB 5; Length 185;
 Best Local Similarity 28.0%; Pred. No. 7.3e-09;
 Matches 53; Conservative 38; Mismatches 76; Indels 22; Gaps 9;

 QY 17 EGCDSLPVISIPVAFGLNATRLG---FSPLESVPFSCIYAGASQFVTMTMLAAG 73
 DB 1 EGAQAAMPALGVISIGLACG---TIGAPYVTPVENGLMSLFYAGSAQFAMALITVQ 56

 QY 74 SSLWIAALTVMAMDMVRHLYGPSLRRIIQRLOKSKTALW----AFGLTDEVFAAATAK 128

Db 57 APVAIAIATVFLINLR--LFILSLHASTYFR----HTSLWNIGMSSILTDITGVLMGE 110
Qy 129 LVNRNRNSENWMTGIAFSSSSWVFGTVIGAFSGSLQGYPAVEAALGFMLPALFMSF 188
Db 111 LAHTD-KVNPWMHGNLNSVAMFVGTVGTALG-GLLPN-PBI-FGLDPAIVGMFIGI 166
Qy 189 LLASFQKQ 197
Db 167 FASQFQMQ 175

RESULT 13

US-10-156-761-10035

; Sequence 10035, Application US/10156761

; Publication No. US20030119018A1

; GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI

; APPLICANT: IKEDA, HARUO

; APPLICANT: ISHIKAWA, JUN

; APPLICANT: HORIKAWA, HIROSHI

; APPLICANT: SHIBA, TADAYOSHI

; APPLICANT: SAKAKI, YOSHIYUKI

; APPLICANT: HATTORI, MASAHIRA

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

; FILE REFERENCE: 249-262

; CURRENT APPLICATION NUMBER: US/10/156,761

; CURRENT FILING DATE: 2002-05-29

; PRIOR APPLICATION NUMBER: JP 2001-204089

; PRIOR FILING DATE: 2001-05-30

; PRIOR APPLICATION NUMBER: JP 2001-272697

; PRIOR FILING DATE: 2001-08-02

; NUMBER OF SEQ ID NOS: 15109

; SEQ ID NO 10035

; LENGTH: 302

; TYPE: PRT

; ORGANISM: Streptomyces avermitilis

US-10-156-761-10035

Query Match 12.3%; Score 153.5; DB 4; Length 302;
Best Local Similarity 27.8%; Pred. No. 7.8e-07;
Matches 59; Conservative 36; Mismatches 96; Indels 21; Gaps 8;

Qy 20 KDSLPIVISYIPVAFAGLGNATRLGFSPLSFVFFSCIIYAGASQFVTAMLAAGSSLWIA 79
Db 72 RDALGVGVAVGLSGFAGFVTSAGSLSLLOTCAUSLLVFTGASQFALVGLAAGNPLAA 131
Qy 80 ALTVMMDVRHVLVGPISRSRI-IQRLQSKTALWAFGLTDEVFAAAATKLVNRNRNRS 138
Db 132 AAGAFFLGVNRFYGLRLSOLLALPRAVRPFAAQW---VIDETTAVALAQPTRRSVR--- 185
Qy 139 NWMIGI---AFSSSSWVFGTVIGAFSGSLQGYPAVEA-ALGFMLPALFMSFL--LAS 192
Db 186 ---IGFTVTGLSLVYLNLTLLIGAVGA---QAIGDTDANGDLDAAGPAVFLALLAPMLK 238
Qy 193 FORKSLCVTAALVG-ALAGVTLSFIPVAILA 223
Db 239 TAARAVAGIAVLGLGLLPVLPAGVPVLA 270

RESULT 14

US-09-738-626-4887

; Sequence 4887, Application US/09738626

; Publication No. US20020197605A1

; GENERAL INFORMATION:

; APPLICANT: NAKAGAWA, SATOSHI

; APPLICANT: MIZOGUCHI, HIROSHI

; APPLICANT: ANDO, SEIKO

; APPLICANT: HAYASHI, MIKIRO

; APPLICANT: OCHIAI, KEIKO

; APPLICANT: YOKOI, HARUHIKO

; APPLICANT: TATEISHI, NAKO

; APPLICANT: SENOH, AKIHIRO

; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4887
; LENGTH: 324
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4887

Query Match 8.4%; Score 105; DB 3; Length 324;
Best Local Similarity 27.4%; Pred. No. 0.066;
Matches 48; Conservative 26; Mismatches 75; Indels 26; Gaps 8;

Qy 69 MLAGSSLIWIAALTV-NAMDVHVLVGPISRSRI-IQRLQSKTALWAFGLTDEVFAA--- 124
Db 21 MLNNGALVGLIALCVGLFIATPHELTIPNL---INIGIQSATVAILAFGTFVITAGID 77
Qy 125 -ATAKLVRNRRNSENWMTGIAFSSSSWVFGTVIGAFSGSLQGYPAVEAALGFMLPA 183
Db 78 LSVGSVAALGAMTSAYFFAEVGLPGWITLLIGLFIGLLAGA-----ISGISIAYG-KLPA 131
Qy 184 LFMSFLLASFQKQSLCV-----TAALVGLAGVTLFSPVAVI-----LAGIYC 227
Db 132 FIATLAWMSIARGITLVISQSPISPAVNAL-GRTYFGIPMPILMMALAGIYC 185

RESULT 15

US-09-738-626-6653

; Sequence 6653, Application US/09738626

; Publication No. US20020197605A1

; GENERAL INFORMATION:

; APPLICANT: NAKAGAWA, SATOSHI

; APPLICANT: MIZOGUCHI, HIROSHI

; APPLICANT: ANDO, SEIKO

; APPLICANT: HAYASHI, MIKIRO

; APPLICANT: OCHIAI, KEIKO

; APPLICANT: YOKOI, HARUHIKO

; APPLICANT: TATEISHI, NAKO

; APPLICANT: SENOH, AKIHIRO

; APPLICANT: IKEDA, MASATO

; APPLICANT: OZAKI, AKIO

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

; FILE REFERENCE: 249-125

; CURRENT APPLICATION NUMBER: US/09/738,626

; CURRENT FILING DATE: 2000-12-18

; PRIOR APPLICATION NUMBER: JP 99/377484

; PRIOR FILING DATE: 1999-12-16

; PRIOR APPLICATION NUMBER: JP 00/159162

; PRIOR FILING DATE: 2000-04-07

; PRIOR APPLICATION NUMBER: JP 00/280988

; PRIOR FILING DATE: 2000-08-03

; NUMBER OF SEQ ID NOS: 7059

; SOFTWARE: PatentIn ver. 3.0

; SEQ ID NO 6653

; LENGTH: 474

; TYPE: PRT

; ORGANISM: Corynebacterium glutamicum

US-09-738-626-6653

Query Match 8.2%; Score 102.5; DB 3; Length 474;
Best Local Similarity 22.9%; Pred. No. 0.19;
Matches 58; Conservative 35; Mismatches 87; Indels 73; Gaps 12;

```
Qy 20 KDSLPIVISYIPVA-----FAFGLNATRLGSPSPLESVPFSCIIYAGASQFVITA 68
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 60 QDGVLLALGLFLPVSWIVRTMLLVAGFAGAWGMBLGPBK-----FLAVTVAIYNPFVVER 114
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 69 MLAGSSSL-----WIAALTYMAMDVHVLYGFSLSRIIQRLOKSKTALWAFGL--TDEVF 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 115 LLOGHWSLVMAVWLLPLVVALR--RHP-----RWQVV-----AIWAASLTPTGAVV 158
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 123 AAATAKLVRNNRRWSENMMIGIAFSSWSSWV-----EGTVIGA 160
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 159 AAVTGVASSKKRFTTL-----CSFLSWLPWLIIPALLATPTSGGALTFAIRSETYAGTLGT 214
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 161 PSGSGLLQGYPAVEAA--LGFMLPA-LFMSFLLASFORQKSLCVTAALVGLAGV----- 212
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 215 ALGLGGIWNAGAVPASRELGFVAGIILLFALLAGFKNCPPWLLALLAVVGFMGAIQPLWM 274
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 213 -TLESIPVAILAG 224
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 275 PNLFTWTIAYVPG 287
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Search completed: February 15, 2006, 12:29:33
Job time : 158.598 secs

This Page Blank (uspto)

November 2005

Published_Applications Nucleic Acid and Published_Applications Amino Acid database searches now generate two sets of results each. The Published_Applications databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches.

Newly published applications will appear in the Published_Applications_New databases; older published applications make up the Published_Applications_Main databases.

Searches run against Nucleic Acid Published_Applications produce two sets of results, with the extensions **.rnpbm** (Published_Applications_NA_Main) and **.rnpbn** (Published_Applications_NA_New).

Searches run against Amino Acid Published_Applications produce two sets of results, with the extensions **.rapbm** (Published_Applications_AA_Main) and **.rapbn** (Published_Applications_AA_New).

This Page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.
OM protein - protein search, using sw model
Run on: February.15, 2006, 12:25:52 ; Search time 14.4522 Seconds
(without alignments)
240.922 Million cell updates/sec
Title: US-10-073-293A-4
Perfect score: 1248
Sequence: 1 MESSTPQAPGSAFMFGCK.....VCGCLTALIQAFMQGAPDEL 245
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 107799 seqs, 14211699 residues
Total number of hits satisfying chosen parameters: 107799
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA New:
1: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
Red. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	229.5	18.4	230	6	US-10-793-626-268 Sequence 268, App
2	105	8.4	304	7	US-11-082-389-14 Sequence 14, Appl
3	91.5	7.3	432	6	US-10-873-528-21 Sequence 21, Appl
4	91.5	7.3	648	6	US-10-793-626-1060 Sequence 1060, Ap
5	88	7.1	251	7	US-11-054-515-1316 Sequence 1316, Ap
6	87	7.0	472	6	US-10-467-557-2268 Sequence 2268, Ap
7	87	7.0	473	6	US-10-467-557-1874 Sequence 1874, Ap
8	86	6.9	300	7	US-11-098-686-10583 Sequence 10583, A
9	86	6.9	469	7	US-11-195-739-19 Sequence 19, Appl
10	85.5	6.9	453	6	US-10-793-626-3260 Sequence 3260, Ap
11	85.5	6.9	500	7	US-11-012-668-4 Sequence 4, Appli
12	83.5	6.7	508	6	US-10-467-557-7826 Sequence 7826, Ap
13	83	6.7	539	6	US-10-467-557-3146 Sequence 3146, Ap
14	82.5	6.6	475	7	US-11-074-176-336 Sequence 336, App
15	82.5	6.6	488	7	US-11-074-176-126 Sequence 126, App
16	82	6.6	234	7	US-11-072-512-2791 Sequence 2791, Ap
17	82	6.6	693	6	US-10-467-557-6176 Sequence 6176, Ap
18	81.5	6.5	526	7	US-11-094-317-43 Sequence 43, Appl
19	81.5	6.5	539	7	US-11-210-316-26 Sequence 26, Appl
20	81.5	6.5	724	7	US-11-043-889-8 Sequence 8, Appli
21	80.5	6.5	252	7	US-11-054-515-1326 Sequence 1326, Ap
22	80	6.4	300	6	US-10-858-730-109 Sequence 109, App
23	80	6.4	455	7	US-11-098-686-10673 Sequence 10673, A
24	79.5	6.4	463	6	US-10-467-557-6352 Sequence 6352, Ap
25	79.5	6.4	463	6	US-10-467-557-7604 Sequence 7604, Ap

26	79.5	6.4	573	6	US-10-055-877-79 Sequence 79, Appl
27	79.5	6.4	634	7	US-11-169-041-189 Sequence 189, App
28	79.5	6.3	248	7	US-11-054-515-1360 Sequence 1360, Ap
29	78.5	6.3	254	7	US-11-054-515-844 Sequence 844, App
30	78.5	6.3	429	7	US-11-205-109-32 Sequence 32, Appl
31	78.5	6.3	439	7	US-11-082-389-80 Sequence 80, Appl
32	78.5	6.3	564	7	US-11-082-389-78 Sequence 78, Appl
33	78.5	6.3	1275	6	US-10-724-598-49 Sequence 49, Appl
34	78	6.2	529	6	US-10-858-730-104 Sequence 104, App
35	78	6.2	529	6	US-10-858-730-105 Sequence 105, App
36	78	6.2	984	7	US-11-055-822-508 Sequence 508, App
37	78	6.2	984	7	US-11-055-822-594 Sequence 594, App
38	77.5	6.2	249	7	US-11-054-515-1321 Sequence 1321, Ap
39	77.5	6.2	650	6	US-10-873-528-95 Sequence 95, Appl
40	77	6.2	249	7	US-11-054-515-957 Sequence 957, App
41	76.5	6.1	316	7	US-11-072-512-2079 Sequence 2079, Ap
42	76.5	6.1	489	6	US-10-858-730-198 Sequence 198, App
43	76.5	6.1	489	7	US-11-055-822-1152 Sequence 1152, Ap
44	76	6.1	448	6	US-10-873-528-42 Sequence 42, Appl
45	75.5	6.0	254	7	US-11-054-515-1659 Sequence 1659, Ap

ALIGNMENTS

RESULT 1
US-10-793-626-268
; Sequence 268, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: P0348008
; CURRENT APPLICATION NUMBER: US/10793.626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 268
; LENGTH: 230
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-268

Query Match	18.4%	Score	229.5	DB	6	Length	230
Best Local Similarity	28.6%	Pred. No.	2.8e-13				
Matches	65	Conservative	47	Mismatches	102	Indels	13
Gaps	6						
Qy	14	TFMEGCKDSLPIVISIYPVAFAGLNATRLGFPLESVPFSCIIYACASQFVITAMLAAG	73				
Db	5	TFQGVKCEIPTLLTGAGVGLSGFVAVSQNFVLEITLLCLIIYAGAAQFIICTLVIAG	64				
Qy	74	SSLIWALATVMADMVDRHVLVGPISLRRIIQRLOKSKTALWAFGLTDFVFAAATAKLRNN	133				
Db	65	TPISAIVITLIVNSRMFLSLMTLAPNYKYGFNVRVGLGTL-LTDTFFGVAITPYK-G	122				
Qy	134	RRSENMMIGIAFSWSSWVFGTVIGAFSGSLGQYPAVEAALG--FMLPALFMSFLLA	191				
Db	123	EKINDRWLHGLNITAVLFTWVSCVIGAFGE----YISNPDALGLDFAITAMFICLS	177				
Qy	192	SFO--RKQSLCVTAALVAGLAGVTLF--SIPVAILAGIVCGCLTALI	234				
Db	178	QFEGIKKRLRIYIVLVIVCVVMMLLSILPSYLAITLIIAAVAALL	224				
RESULT 2							
US-11-082-389-14							
; Sequence 14, Application US/11082389							
; Publication No. US20050244935A1							

```
; GENERAL INFORMATION:
; APPLICANT: Pompeius, Markus
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberhauer, Gregor
; TITLE OF INVENTION: INVOLVED IN MEMBRANE SYNTHESIS AND MEMBRANE
; TITLE OF INVENTION: TRANSPORT
; FILE REFERENCE: BGI-13ICPCN
; CURRENT APPLICATION NUMBER: US/11/082,389
; CURRENT FILING DATE: 2005-03-16
; PRIORITY APPLICATION NUMBER: US 09/603024
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/141031
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 60/143262
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: US 60/151281
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: DE 19930487.4
; PRIOR FILING DATE: 1999-07-01
; PRIOR APPLICATION NUMBER: DE 19930489.0
; PRIOR FILING DATE: 1999-07-01
; PRIOR APPLICATION NUMBER: DE 19931549.3
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931550.7
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19932134.5
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19941379.7
; PRIOR FILING DATE: 1999-08-31
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 446
; SEQ ID NO 14
; LENGTH: 304
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-11-082-389-14

Query Match      8.4%; Score 105; DB 7; Length 304;
Best Local Similarity 27.4%; Pred. No. 0.028;
Matches 48; Conservative 26; Mismatches 75; Indels 26; Gaps 8;

Qy 69 MLAGSSLWIAALTY-MAMDVHVLVYGPSLSRSRIQRLQSKTALWAFGLTDEVPAA--- 124
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MLNRGALVGLIALCVGLFIATPHELTIPNL---INIGIQSATVAIALAFGMTFVITAGID 57

Qy 125 -ATAKLVRNNRRNSNMIGIAFSSWSWVFGTVIGAFSGSGLLQGYPAVEAALGFMLPA 183
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 58 LSVGSVAALGAMTSAFFAEVGLPGWITLLGLFGLLAGA-----ISGISIAYG-KLPA 111

Qy 184 LFMSFLIASFORKOSLCV-----TAALVGLAGVTLFSPVAI-----LAGIYC 227
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 112 FIATLAMMSIARGITLVISQGSPIPSAPAVNAL-GRTVFGIPMPILMWALAGIYC 165

RESULT 3
US-10-873-528-21
; Sequence 21, Application US/10873528
; Publication No. US20050276814A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129WO
; CURRENT APPLICATION NUMBER: US/10/873,528
; CURRENT FILING DATE: 2004-06-23
; PRIOR APPLICATION NUMBER: US/09/769,787
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
```

```
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-873-528-21

Query Match      7.3%; Score 91.5; DB 6; Length 432;
Best Local Similarity 22.9%; Pred. No. 0.61;
Matches 61; Conservative 39; Mismatches 83; Indels 83; Gaps 15;

Qy 30 IPVAFAPGLNATRLGFSPLSVFSCIIYAGASQFVTITMLAAGSS---LWIAALTWAM 86
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 171 VPQAIARQFEA---MIPAFVIFLSSMIV-----YILAKSLTNGGTFIEMIYSAIQVPLQ 221

Qy 87 DVRHVLVGPSLSRSRIQRLQSKTALWAFGLTDS--VFAAATAKLVRN----- 132
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 222 GLTGSGLYGAIGIAFFI-----SFLWWFGVHGQSVVGVVTTALLSLNLDANKAMLASAN 274

Qy 133 -----NRRWSENWMI-----GIAFSSWSWVFGTVIGAFSGSGLLQGYPAV----- 173
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 275 LSLNGAHIVTQQFLDSFLILSGSGITFGLVWAMLFAAKSKQYQALGKVAAPAIENVNE 334

Qy 174 EAALGP---MLPALFMSFLIASFORKOSLCVTAALV--GALA-----GVTL-FSIPV 219
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 335 PVVGFPIVMNPVMFVPFILVP-----VLAIVIVGAIGATGMQPFSGVTLPWSTP- 385

Qy 220 AILAGIVCGCLTALIQAFMQGAPDEL 245
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 386 ALLSGFLVG-----GWQGVITQL 403

RESULT 4
US-10-793-626-1060
; Sequence 1060, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1060
; LENGTH: 648
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-1060

Query Match      7.3%; Score 91.5; DB 6; Length 648;
Best Local Similarity 21.0%; Pred. No. 0.94;
Matches 63; Conservative 34; Mismatches 98; Indels 105; Gaps 15;

Qy 9 APGSATFMEGCKSLPIVISYIPVAFAPGLNATRLGFSPLSVFSCIIYA----- 59
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 156 APLAGEFSPG-----PGVNTYLIATQISGISGLMTGIN-----FFVTILRCKTPTMKFMQ 205

Qy 60 ----GASQFVITAMLAAGSSIIWIAALTVAMDVHVLVGPSLSRSRIQRLQSKTALW-- 113
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 206 MPMSFVTFITTLIVILAFVPFVTTVALMTAD---RIFG-----TQFTVANGGMPMLWAN 258

Qy 114 -----AFGLTDEVFAAATAKLVRNNRRWSENWMI-----GIAFSSWSWV 153
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 259 PFVWVGHPHYVIVILPAFGMYSEIIPTFARK-----RLFHQSMIWTAGIAFLSFLVWV 313
```


Db 255 KAQAIVATDSNTQLSIAFVFNCLLLTLGAALFFGTTKEDLGGFYDLYLALKTEPALGA 314
Qy 176 ALGFMPLPALFMSFLLASFORQSLCVTAALVG--ALAGVTLFSIP-----VAILA 223
Db 315 TLGGIMSTLFAVALLAS---QNSTITGTLAGQIVMEGFLKLSIPNWLRLIITRSLAIVP 371
Qy 224 GIVC 227
Db 372 VIIC 375

RESULT 11

US-11-012-668-4
; Sequence 4, Application US/11012668
; Publication No. US20050010512A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; APPLICANT: Rafalski, J. Antoni
; TITLE OF INVENTION: Nitrogen Transport Metabolism
; FILE REFERENCE: BB-1210
; CURRENT APPLICATION NUMBER: US/11/012.668
; PRIOR FILING DATE: 2004-12-15
; PRIOR APPLICATION NUMBER: 60/098,248
; PRIOR FILING DATE: 28 August 1998
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 4
; LENGTH: 500
; TYPE: PRT
; ORGANISM: Glycine max
US-11-012-668-4

Query Match 6.9%; Score 85.5; DB 7; Length 500;
Best Local Similarity 23.1%; Pred. No. 2.4;
Matches 65; Conservative 30; Mismatches 89; Indels 97; Gaps 16;

Qy 29 YIVAFAPAGLNATRLGP-----SPLESVFFSCIIYAGASQFVITAM-LAAGSSWL 77
Db 97 YLFGFAFAGFSPNGPIGKHFFGLKDIPSSSYDYSVFLYQWA--FAIAAAGITSGS-- 151
Qy 78 IALITWAMDVRHVLGSPRSRIIQLQSKTALW-----AFGLTDEVFAA----- 124
Db 152 IABTOP---VAVLIYSSFLTGTVPYVSH-----WFWSPDGWASAFKITDRLFTSGVID 203
Qy 125 -ATAKLVRNRRNRSNWMIGIAFSSWSWVGTGVIAGFSGG---LLOGYPVAEALG-F 179
Db 204 FAGSGVH-----WVGGIAGLWGLLEGPRMGRFDHAGRAVALRGHSLSVLVLTGF 254
Qy 180 ML-----PALFMSLL-----ASFQKQSLCVTAALVGLAGV-TLFSIPV-- 219
Db 255 LLWFGWYGFNPGSFNKLITLYGNSGNYGQSAVGRTAVTTTLAGSTAALTTLFGKRVIS 314
Qy 220 -----AIIAG-----IIVCGCLTALI 234
Db 315 GHNVTVDVNCNLLGGLGFAAITAGCSVVEPMAIIVCGFVASIV 355

RESULT 12

US-10-467-657-7826
; Sequence 7826, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12

; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 7826
; LENGTH: 508
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-7826

Query Match 6.7%; Score 83.5; DB 6; Length 508;
Best Local Similarity 18.9%; Pred. No. 3.6;
Matches 63; Conservative 39; Mismatches 108; Indels 123; Gaps 11;

Qy 25 IIVSYIPVAFAGLNATRLGFSPL-ESVFFSCIIYAGASQFVITAMLA-----G 73
Db 129 IILFFFTIYCAGIVAGATLFSQSLFEGMTYNQAMWLGAGATIAVTLGGFLAVSWTDTLQ 188
Qy 74 SSLWIAALTVMAMDVRHVLGSPRSRIIQLQSKS-----KTALWAF 115
Db 189 ASLMIFALITLTPVMVYVGLGGABQMSAAIQSVAAGTKGYGSLFAGTTVIGIITAAWGL 248
Qy 116 G-----LTDEVFAAATAKLVRNRRNRSNWM-----IGIAF-----SSW 149
Db 249 GYFGQPHILARFMAESAASKLSVSARRIGMTWMALCLAGAVAGVFGIAYFGANPDKUSSM 308
Qy 150 SS-----WVFGTVIGAFSGSGLLQGYPAVEAALGFM-----PALF 185
Db 309 SGNHERIFIALSTLLFPNPMIAGIILSAI-----LAAVMTLSCLLVCSAITEDFY 360
Qy 186 MSFLLASFORQSLCVTAALVGLAGVTLFSI-----PVAL- 221
Db 361 KGFLRKVAQOQSELVWVGRMLVLAIVISILIASDPNSKVLGLVSYAWAGFAAGFPIVIL 420
Qy 222 -----LAGIVCGCLTALIOAFWQGP 242
Db 421 SVLWKEITAYGALSGNVAGASTVTVVWAEWVKP 453

RESULT 13

US-10-467-657-3146
; Sequence 3146, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 3146
; LENGTH: 539
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-3146

Query Match 6.7%; Score 83; DB 6; Length 539;
Best Local Similarity 22.1%; Pred. No. 4.3;
Matches 60; Conservative 41; Mismatches 100; Indels 70; Gaps 15;

Qy 9 AFGSATFMGCKDSDLIVISYIPVAF--AFGLNATRLGSPLESVFFSCIIYAGASQFVI 66
Db 240 AMGMLSHLDREKHKTYPVFLLLGIAYCASIGGLGTIVGSP-----NLIAAKALMLDF 292
Qy 67 TAWLAAGSSIIWIAALTVMAMDVRHVLGSPRSRIIQLQK-----SKTA 111
Db 293 VGMKLGLPMLLILPLMLLSL--YVILKPNLNERVEIKAESIPWTLHRVALLIFLATAA 351

```
Qy 112 LW-----AFGLT--DEVPA--AATAKLVNRNRWSE-----NWMI-----GIAFS- 147
Db 352 AWIFGSKITAFGINSNEDTVIALSAAVAVVVGVAQWKEVARNTDVGVLMLFGGGISLST 411
Qy 148 ----SWSSWVGTVIGA-FSGSGLLQGYPAVEAALGFMLPALFMSFLLASFQKQSLCVT 202
Db 412 LLKTSGASEALGQQVAATFSGA-----PAFLVILIVAFAFIIFLTPTSNITAS---A 459
Qy 203 AALVAGALAGVTL-PSIPVAILA-----GIVCGC 229
Db 460 ALLVPIFSGIAMQGLPEQVLVFIIGIAASC 490
```

RESULT 14

```
US-11-074-176-336
; Sequence 336, Application US/11074176
; Publication No. US20050250135A1
; GENERAL INFORMATION:
; APPLICANT: Klaenhammer, Todd R.
; APPLICANT: Russell, William M.
; APPLICANT: Altermann, Eric
; APPLICANT: McAuliffe, Olivia
; APPLICANT: Perill, Andrea Azcarate
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding
; TITLE OF INVENTION: Stress-Related Proteins and Uses Therefore
; FILE REFERENCE: 5051-694
; CURRENT APPLICATION NUMBER: US/11/074.176
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: 60/551,161
; PRIOR FILING DATE: 2004-03-08
; NUMBER OF SEQ ID NOS: 381
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 336
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Lactobacillus acidophilus
US-11-074-176-336
```

```
Query Match 6.6%; Score 82.5; DB 7; Length 475;
Best Local Similarity 26.1%; Pred. No. 4.1;
Matches 49; Conservative 25; Mismatches 75; Indels 39; Gaps 8;

Qy 70 LAAG-SSLWIAALTVMAMDVHVLVYGPSLRSRIQRLQKSKTALWAFGLTDEVFAAATAK 128
Db 33 LALGVGTIVSASIFTLPGEVAALHTGPAVAISFI-----LAAVAAGLVAFAYAEAAA 85
Qy 129 LVNRNRWSENWMIIGIAFSSWSSWVFG-----TVIGAFSGSGLLQGYPAVEAALGFML 181
Db 86 MPFAGSAYS--W-INNVFGEFGWVAGWALLAEYFIALAFVGSGLSANFRALIAPLGIKL 142
Qy 182 PAL-----FMSFLLASFQKQSLCVTAAALVAGALAGVTLFSIPVAILAGI 225
Db 143 PAALSNAFGTGGVIDIISILISILLVALLSHGVSRARVENALVVLKV-----LAILFFI 198
Qy 226 VCGCLTAL 233
Db 199 VVG-LTAI 205
```

RESULT 15

```
US-11-074-176-126
; Sequence 126, Application US/11074176
; Publication No. US20050250135A1
; GENERAL INFORMATION:
; APPLICANT: Klaenhammer, Todd R.
; APPLICANT: Russell, William M.
; APPLICANT: Altermann, Eric
; APPLICANT: McAuliffe, Olivia
; APPLICANT: Perill, Andrea Azcarate
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding
; TITLE OF INVENTION: Stress-Related Proteins and Uses Therefore
; FILE REFERENCE: 5051-694
; CURRENT APPLICATION NUMBER: US/11/074.176
```


```
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: 60/551,161
; PRIOR FILING DATE: 2004-03-08
; NUMBER OF SEQ ID NOS: 381
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 126
; LENGTH: 488
; TYPE: PRT
; ORGANISM: Lactobacillus acidophilus
US-11-074-176-126

Query Match 6.6%; Score 82.5; DB 7; Length 488;
Best Local Similarity 26.1%; Pred. No. 4.2;
Matches 49; Conservative 25; Mismatches 75; Indels 39; Gaps 8;

Qy 70 LAAG-SSLWIAALTVMAMDVHVLVYGPSLRSRIQRLQKSKTALWAFGLTDEVFAAATAK 128
Db 46 LALGVGTIVSASIFTLPGEVAALHTGPAVAISFI-----LAAVAAGLVAFAYAEAAA 98
Qy 129 LVNRNRWSENWMIIGIAFSSWSSWVFG-----TVIGAFSGSGLLQGYPAVEAALGFML 181
Db 99 MPFAGSAYS--W-INNVFGEFGWVAGWALLAEYFIALAFVGSGLSANFRALIAPLGIKL 155
Qy 182 PAL-----FMSFLLASFQKQSLCVTAAALVAGALAGVTLFSIPVAILAGI 225
Db 156 PAALSNAFGTGGVIDIISILISILLVALLSHGVSRARVENALVVLKV-----LAILFFI 211
Qy 226 VCGCLTAL 233
Db 212 VVG-LTAI 218
```

Search completed: February 15, 2006, 12:30:00
Job time : 15.4522 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - nucleic search, using  model

Run on: February 16, 2006, 17:05:59 ; Search time 6212.4 Seconds
(without alignments)
2241.749 Million cell updates/sec

Title: US-10-073-293A-4
Perfect score: 1248
Sequence: 1 MESPTQPAPGSAFTMEGCK.....VCGCLTALIQAFWQGPDEL 245

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 5883141 seqs, 20421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/abs/ABSSWEB_spool/US10073293/runat_15022006_120734_963/app_query.fasta.1
-DB=genEmbl -QMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs04
-USER=US10073293 @CGN 1.1 5142 @runat_15022006_120734_963 -NCPU=6 -ICPU=3
-NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl: *
1: gb_ba: *
2: gb_in: *
3: gb_env: *
4: gb_ov: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pr: *
9: gb_ro: *
10: gb_sts: *
11: gb_sy: *
12: gb_un: *
13: gb_vi: *
14: gb_htg: *
15: gb_pl: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1248	100.0	738	6	BD177948 Process f
2	1248	100.0	738	6	CS078097 Sequence
3	1248	100.0	738	6	CS078113 Sequence

4	1248	100.0	738	6	AX534668
5	1248	100.0	19150	1	D90891
6	1248	100.0	110000	1	U00096_27
7	1248	100.0	110000	1	U00096_28
8	1247	99.9	110000	1	AE005174_35
9	1247	99.9	110000	1	AE005174_36
10	1247	99.9	110000	1	BA000007_35
11	1247	99.9	305325	1	AE016765
12	1243	99.6	110000	1	AE005674_27
13	1243	99.6	290380	1	AE016987
14	1013.5	81.2	756	6	AR386287
15	871.5	69.8	10733	1	AE013695
16	871.5	69.8	220050	1	AJ414156
17	871.5	69.8	290002	1	AE017129
18	870.5	69.8	110000	1	EX936398_10
19	845.5	67.7	110000	1	EX950851_39
20	841.5	67.4	243072	6	AX770904
21	841.5	67.4	349318	1	EX571863
22	834	66.8	3619	1	EC410307
23	592	47.4	801	6	AR319849
24	431	34.5	190822	14	AC051613
25	385	30.8	348624	1	EX640441
26	385	30.8	348666	1	EX640426
27	362	29.0	343243	1	EX640414
28	347	27.8	202872	14	AC016160
29	317.5	25.4	13099	1	AE000982
30	308.5	24.7	6002	1	AF118548
31	308.5	24.7	110000	1	AE014291_17
32	308.5	24.7	110000	1	AE017223_17
33	308.5	24.7	110000	1	AE017223_18
34	301.5	24.2	349648	1	EX957220
35	300.5	24.1	110000	1	AP008226_02
36	300.5	24.1	247910	1	AE017307
37	287	23.0	110000	1	BA000004_30
38	281	22.5	110000	1	BA000043_06
39	278.5	22.3	110000	1	AP006840_04
40	273	21.9	110000	1	CP000090_05
41	269	21.6	110000	1	AE017225_16
42	269	21.6	110000	1	AE017334_16
43	269	21.6	110000	1	AE017355_16
44	269	21.6	230525	1	AE017029
45	268.5	21.5	110000	1	AE017333_10

ALIGNMENTS

BD177948	Process for producing L-amino acid using escherichia.	738 bp	DNA	linear	PAT 16-APR-2003
BD177948	Process for producing L-amino acid using escherichia.				
BD177948	GI:30015211				
JP 2002300874-A/3.	Escherichia coli				
OS	Escherichia coli				
PN	JP 2002300874-A/3				
PD	15-OCT-2002				
PR	13-FEB-2002				
PR	13-FEB-2001				
PR	13-FEB-2001				
26-FEB-2001					
28-JUN-2001					
PI	EKATERINA ALEKSANDROVNA TABOLINA, KONSTANTIN VYACHESLAVOVICH				
PI	RYBAK,				
PI	EVGENI MOISEVICH KHOURGES, ELVIRA BORISOVNA VOROSHILOVA, PI				

```
MIKHAIL MAROVICH GUSYATINER
PC C12N1/21,C12P13/04,C12P13/06,C12P13/08,C12P13/12, PC
C12P13/24//
PC (C12N1/21,C12R1:19),(C12P13/04,C12R1:19),(C12P13/06,C12R1:19),
PC (C12P13/08,C12R1:19),(C12P13/08,C12R1:19),(C12P13/12,C12R1:19), PC
(C12P13/24,C12R1:19)
CC Process for producing L-amino acid using escherichia FH Key
FT CDS Location/Qualifiers
FEATURES
    source      1..738
                /organism="Escherichia coli"
                /mol_type="genomic DNA"
                /db_xref="taxon:562"
ORIGIN
Alignment Scores:
Pred. No.:      5,54e-107      Length:      738
Score:          1248.00      Matches:      245
Percent Similarity: 100.0%      Conservative: 0
Best Local Similarity: 100.0%      Mismatches: 0
Query Match:     100.0%      Indels:      0
DB:              6           Gaps:          0
US-10-073-293A-4 (1-245) x BD177948 (1-738)
Qy 1 MetGluSerProThrProGlnProAlaProGlySerAlaThrPheMetGluGlyCysLys 20
Db 1 ATGGAAGCCCTACTCCACAGCCTGCTCCTGGTTCGGCGACCTTCATGGAAGGATGCAAA 60
Qy 21 AspSerLeuProIleValIleSerTyrIleProValAlaPheAlaPheGlyLeuAsnAla 40
Db 61 GACAGTTTACCAGATTGTTATTAGTTATATTCCGGTGGCCTTTGGCTTCGATGCG 120
Qy 41 ThrArgLeuGlyPheSerProLeuGluSerValPhePheSerCysIleIleTyrAlaGly 60
Db 121 ACCCGTCTGGGATTCCTCCCTCGAAAGCGTTTTTTTCTCTGCATCATTTATGCGGC 180
Qy 61 AlaSerGlnPheValIleThrAlaMetLeuAlaGlySerSerLeuTrpIleAlaAla 80
Db 181 GCGAGCCAGTTCGTCAATACCGCATGCTGGCAGCGCGAGTAGTTTGGATTGCTGCA 240
Qy 81 LeuThrValMetAlaMetAspValArgHisValLeuTyrGlyProSerLeuArgSerArg 100
Db 241 CTGACCGTCATGCGCAATGGATGTTGCGCATGTGTTGATGCGCCGTCCTACTGCGTAGCCGT 300
Qy 101 IleIleGlnArgLeuGlnLysSerLysThrAlaLeuTrpAlaPheGlyLeuThrAspGlu 120
Db 301 ATTATTACGCGTCTGCAAAATTCGAAACCGCCCTGTGGGCTTTGGCTGACGGATGAG 360
Qy 121 ValPheAlaAlaThrAlaLysLeuValArgAsnAsnArgArgTfTrpSerGluAsnTrp 140
Db 361 GTTTTTCGCGCGCAACCGCAAACTGGTAGCAATAATCGCGCTGCGAGCGAGAACTGG 420
Qy 141 MetIleGlyIleAlaPheSerSerTrpSerSerTrpValPheGlyThrValIleGlyAla 160
Db 421 ATGATCGCGATGCGCTTCAGTTCATGGTCATCGTGGGTATTGGTACGGTAATAGGGCA 480
Qy 161 PheSerGlySerGlyLeuLeuGlnGlyTyrProAlaValGluAlaAlaLeuGlyPheMet 180
Db 481 TTCTCCGCGAGCGGTTCGTCGAAGTTATCCCGCGTTGAACTGTCATTAGTTTATG 540
Qy 181 LeuProAlaLeuPheMetSerPheLeuLeuAlaSerPheGlnArgLysGlnSerLeuCys 200
Db 541 CTTCCGGCACCTTTATGAGTTTCCTGCTCGCCTCTTCCAGCGCAAAACAATCTCTTGC 600
Qy 201 ValThrAlaAlaLeuValGlyAlaLeuAlaGlyValThrLeuPheSerIleProValAla 220
Db 601 GTTACCGCAGCGTTAGTTGGTCCCTTCGAGCGTAACGCTATTTCCTATTCCTCCGTCGCC 660
Qy 221 IleLeuAlaGlyIleValCysGlyCysLeuThrAlaLeuIleGlnAlaPheTrpGlnGly 240
|||||
```

```
661 ATTCTGCGAGGCAATTGCTGTGGTGCTGCTCACTCGTTAATCCAGGCAATCTCGCAAGGA 720
Qy 241 AlaProAspGluLeu 245
|||||
Db 721 GCGCCCGATGAGCTA 735
RESULT 2
CS078097 738 bp DNA linear PAT 06-MAY-2005
LOCUS Sequence 3 from Patent EP1526179.
DEFINITION CS078097
ACCESSION CS078097
VERSION CS078097.1 GI:63093065
KEYWORDS
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
1 Tabolina,E.A., Rybak,K.V., Khourges,E.M., Voroshilova,E.B. and
Gusyatiner,M.M.
TITLE Method for producing l-amino acid using bacteria belonging to the
Genus Escherichia
JOURNAL Patent: EP 1526179-A 3 27-APR-2005;
Ajinomoto Co., Inc. (JP)
FEATURES
    source      1..738
                /organism="Escherichia coli"
                /mol_type="unassigned DNA"
                /db_xref="taxon:562"
    CDS         1..738
                /note="unnamed protein product"
                /codon_start=1
                /transl_table=11
                /protein_id="CAI94223.1"
                /db_xref="GI:63093066"
                /translation="MESPTQPAPGSATFMEGCKDSLPIVISIYPVAFGLNATRLG
                FSPLESYFFCIIYVAGASQFVITAMLAAGSSLWIAALTVMAMDVVRHLYGSPLSRRII
                ORLQKSTALWAFGLTDEVFAAATAKLVRNNRWSNNMIGIAFSSWSWVGTVIGA
                FSGSLGQGPVAVEMALGFMPLPALFMSFLASFORKQSLCVTAALVGLAGLVTFSP
                VALLAGIVCGCLTALIQAFQWQAPDEL"
    ORIGIN
Alignment Scores:
Pred. No.:      5,54e-107      Length:      738
Score:          1248.00      Matches:      245
Percent Similarity: 100.0%      Conservative: 0
Best Local Similarity: 100.0%      Mismatches: 0
Query Match:     100.0%      Indels:      0
DB:              6           Gaps:          0
US-10-073-293A-4 (1-245) x CS078097 (1-738)
Qy 1 MetGluSerProThrProGlnProAlaProGlySerAlaThrPheMetGluGlyCysLys 20
Db 1 ATGGAAGCCCTACTCCACAGCCTGCTCCTGGTTCGGCGACCTTCATGGAAGGATGCAAA 60
Qy 21 AspSerLeuProIleValIleSerTyrIleProValAlaPheAlaPheGlyLeuAsnAla 40
Db 61 GACAGTTTACCAGATTGTTATTAGTTATATTCCGGTGGCCTTTGGCTTCGATGCG 120
Qy 41 ThrArgLeuGlyPheSerProLeuGluSerValPhePheSerCysIleIleTyrAlaGly 60
Db 121 ACCCGTCTGGGATTCCTCCCTCGAAAGCGTTTTTTTCTCTGCATCATTTATGCGGC 180
Qy 61 AlaSerGlnPheValIleThrAlaMetLeuAlaGlySerSerLeuTrpIleAlaAla 80
Db 181 GCGAGCCAGTTCGTCAATACCGCATGCTGGCAGCGCGAGTAGTTTGGATTGCTGCA 240
Qy 81 LeuThrValMetAlaMetAspValArgHisValLeuTyrGlyProSerLeuArgSerArg 100
Db 241 CTGACCGTCATGCGCAATGGATGTTGCGCATGTGTTGATGCGCCGTCCTACTGCGTAGCCGT 300
Qy 101 IleIleGlnArgLeuGlnLysSerLysThrAlaLeuTrpAlaPheGlyLeuThrAspGlu 120
Db 301 ATTATTACGCGTCTGCAAAATTCGAAACCGCCCTGTGGGCTTTGGCTGACGGATGAG 360
Qy 121 ValPheAlaAlaThrAlaLysLeuValArgAsnAsnArgArgTfTrpSerGluAsnTrp 140
Db 361 GTTTTTCGCGCGCAACCGCAAACTGGTAGCAATAATCGCGCTGCGAGCGAGAACTGG 420
Qy 141 MetIleGlyIleAlaPheSerSerTrpSerSerTrpValPheGlyThrValIleGlyAla 160
Db 421 ATGATCGCGATGCGCTTCAGTTCATGGTCATCGTGGGTATTGGTACGGTAATAGGGCA 480
Qy 161 PheSerGlySerGlyLeuLeuGlnGlyTyrProAlaValGluAlaAlaLeuGlyPheMet 180
Db 481 TTCTCCGCGAGCGGTTCGTCGAAGTTATCCCGCGTTGAACTGTCATTAGTTTATG 540
Qy 181 LeuProAlaLeuPheMetSerPheLeuLeuAlaSerPheGlnArgLysGlnSerLeuCys 200
Db 541 CTTCCGGCACCTTTATGAGTTTCCTGCTCGCCTCTTCCAGCGCAAAACAATCTCTTGC 600
Qy 201 ValThrAlaAlaLeuValGlyAlaLeuAlaGlyValThrLeuPheSerIleProValAla 220
Db 601 GTTACCGCAGCGTTAGTTGGTCCCTTCGAGCGTAACGCTATTTCCTATTCCTCCGTCGCC 660
Qy 221 IleLeuAlaGlyIleValCysGlyCysLeuThrAlaLeuIleGlnAlaPheTrpGlnGly 240
|||||
```

```
Db 301 ATTATTCAGCGCTCTGCAAAAATCGAAAAACCCGCTGTGGCGTTTGGCTGACGGATGAG 360
Qy 121 ValPheAlaAlaAlaThrAlaLysLeuValArgAsnAenArgTTPSerGluAsnTTP 140
Db 361 GTTTTGGCGCGCAACCGCAAACTGGTAGCAATATATCCGCTGGAGCGAAGTGG 420
Qy 141 MetIleGlyIleAlaPheSerSerTTPSerTTPValPheGlyThrValIleGlyAla 160
Db 421 ATGATCGGCATTCGCTTCAGTTTCATGGTTCATCGTGGGTATTGGTACGGTAATAGGGCA 480
Qy 161 PheSerGlySerGlyLeuLeuGlnGlyTTPProAlaValGluAlaAlaLeuGlyPheMet 180
Db 481 TTCTCCGCGCAGCGCTTGTGCAAGGTATATCCCGCTGAAGTGCATTAGGTTTATG 540
Qy 181 LeuProAlaLeuPheMetSerPheLeuLeuAlaSerPheGlnArgLysGlnSerLeuCys 200
Db 541 CTTCGCGCACTCTTTATGAGTTTCTGCTCGCCTCTTTCAGCGCAACAACTCTTTTTC 600
Qy 201 ValThrAlaAlaLeuValGlyAlaLeuAlaGlyValThrLeuPheSerIleProValAla 220
Db 601 GTTACCGCAGCGTTAGTTGGTGCCCTTGCAGGCGTAACGCTATTCTTATTCCCGTCGCC 660
Qy 221 IleLeuAlaGlyIleValCysGlyCysLeuThrAlaLeuIleGlnAlaPheTTPGlnGly 240
Db 661 ATTCGCGCAGCATTTGCTGTGGCTGCCTCACTCGTTAATCCAGGCATTCTGCAAGGA 720
Qy 241 AlaProAspGluLeu 245
Db 721 GCGCCCGATGAGCTA 735
```

```
RESULT 3
CS078113 LOCUS CS078113 738 bp DNA linear PAT 06-MAY-2005
DEFINITION Sequence 3 from Patent EP1526181.
ACCESSION CS078113
VERSION CS078113.1 GI:63093081
```

```
KEYWORDS Escherichia coli
SOURCE Escherichia coli
ORGANISM Escherichia coli
REFERENCE 1
AUTHORS Tabolina,E.A., Rybak,K.V., Khourges,E.M., Voroshilova,E.B. and
Gusyatiner,M.M.
TITLE Method for producing l-amino acid using bacteria belonging to the
Genus Escherichia
JOURNAL Patent: EP 1526181-A 3 27-APR-2005;
Ajinomoto Co., Inc. (JP)
```

```
FEATURES
source 1..738
location/Qualifiers
1..738
/organism="Escherichia coli"
/mol_type="unassigned DNA"
/db_xref="taxon:562"
1..738
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAI94227.1"
/db_xref="GI:63093082"
/translation="MESPTQPAPFGSATFMGCKDSLPIVSIYPVAFGLNATRLG
FSPLESVFFSCIIYAGASQFVITAMLAAGSLMTAALTVMAMDVRHVLVYGPSLRSRII
QRLOKSTALWAFGLTDEVFAAATAKLVRRNRWSENWMIQIAFSSWSWVFGTVIGA
FSGSLGLQGPVAVEAALGFMLPALFMSFLASFORQSCLVCTAALVAGALAGVTLFSP
VALLIAGVCCITLTIQAFWQGPDEL"
```

ORIGIN

```
Alignment Scores:
Pred. No.: 5,54e-107 Length: 738
Score: 1248.00 Matches: 245
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
```

```
DB: 6 Gaps: 0
US-10-073-293A-4 (1-245) x CS078113 (1-738)
Qy 1 MetGluSerProThrProGlnProAlaProGlySerAlaThrPheMetGluGlyCysLys 20
Db 1 ATGAAAGCCCTACTCCACAGCCTGCTCTCGGTTCGGCGACCTTCATGGAAGGATGCAAA 60
Qy 21 AspSerLeuProIleValIleSerTyrIleProValAlaPheAlaPheGlyLeuAsnAla 40
Db 61 GACAGTTTACCGATTGTTATTAGTTATTATCCGTTGGCTTGGTTCGGTCTGAATGCG 120
Qy 41 ThrArgLeuGlyPheSerProLeuGluSerValPhePheSerCysIleIleTyrAlaGly 60
Db 121 ACCCGTCTGGGATTCCTCCCTCGAAAGCGCTTTTTCCTCTCGCATCATTTATGCGAGC 180
Qy 61 AlaSerGlnPheValIleThrAlaMetLeuAlaAlaGlySerSerLeuTTPIleAlaAla 80
Db 181 GCGAGCCAGTTTCGTCATTACCGGATGCTGGCAGCGGAGTAGTTTGTGATTTGCTGCA 240
Qy 81 LeuThrValMetAlaMetAspValArgHisValLeuTyrGlyProSerLeuArgSerArg 100
Db 241 CTGACCGTCATGGCAATGGATGTTCCGCATGTGTGTATGCCCGTCACTGGCGTAGCCGT 300
Qy 101 IleIleGlnArgLeuGlnLysSerLysThrAlaLeuTTPAlaPheGlyLeuThrAspGlu 120
Db 301 ATTATTGAGCGCTCTGCAAAAATCGAAAAACCGCCCTGTGGCGTTTGGCTGACGATGAG 360
Qy 121 ValPheAlaAlaThrAlaLysLeuValArgAsnAenArgTTPSerGluAsnTTP 140
Db 361 GTTTTGGCGCGCAACCGCAAACTGGTAGCAATATATCCGCTGGAGCGAAGTGG 420
Qy 141 MetIleGlyIleAlaPheSerSerTTPSerTTPValPheGlyThrValIleGlyAla 160
Db 421 ATGATCGGCATTTGCTTCAGTTTCATGGTTCATCGGTGATTTGGTACGGTAATAGGGCA 480
Qy 161 PheSerGlySerGlyLeuLeuGlnGlyTTPProAlaValGluAlaAlaLeuGlyPheMet 180
Db 481 TTCTCCGCGCAGCGCTTGTGCAAGGTATATCCCGCTTCCTTCAGCGCAACAACTCTTTTTC 540
Qy 181 LeuProAlaLeuPheMetSerPheLeuLeuAlaSerPheGlnArgLysGlnSerLeuCys 200
Db 541 CTTCGCGCACTCTTTATGAGTTTCTGCTCGCCTCTTTCAGCGCAACAACTCTTTTTC 600
Qy 201 ValThrAlaAlaLeuValGlyAlaLeuAlaGlyValThrLeuPheSerIleProValAla 220
Db 601 GTTACCGCAGCGTTAGTTGGTGCCCTTGCAGGCGTAACGCTATTCTTATTCCCGTCGCC 660
Qy 221 IleLeuAlaGlyIleValCysGlyCysLeuThrAlaLeuIleGlnAlaPheTTPGlnGly 240
Db 661 ATTCGCGCAGCATTTGCTGTGGCTGCCTCACTCGTTAATCCAGGCATTCTGCAAGGA 720
Qy 241 AlaProAspGluLeu 245
Db 721 GCGCCCGATGAGCTA 735
```

RESULT 4

```
AX534668 LOCUS AX534668 738 bp DNA linear PAT 22-NOV-2002
DEFINITION Sequence 3 from Patent EP1239041.
ACCESSION AX534668
VERSION AX534668.1 GI:25261073
KEYWORDS Escherichia coli
SOURCE Escherichia coli
ORGANISM Escherichia coli
REFERENCE 1
AUTHORS Tabolina,E.A., Rybak,K.V., Khourges,E.M., Voroshilova,E.B. and
Gusyatiner,M.M.
TITLE Method for producing l-amino acid using bacteria belonging to the
genus escherichia
JOURNAL Patent: EP 1239041-A 3 11-SEP-2002;
```

FEATURES		Ajinomoto Co., Inc. (JP)	
source	1..738	Location/Qualifiers	
	/organism="Escherichia coli"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:562"		
CDS	1..738		
	/note="unnamed protein product"		
	/codon_start=1		
	/transl_table=11		
	/protein_id="CAD57225.1"		
	/db_xref="GI:25261074"		
	/translation="MSESTPQAPGSAFTMEGCKDSLPVIVSYIPVAPAFGLNATRLG		
	FSPLESVPFCIIIVAGASQFVITAMLAAGSSLTWALTVMAMDVRHLVGPSTLSRII		
	ORLQSKTALWAFGLTDEVFPAATAKLVRRNRNSENWMIIGIAFPSSSWVFGTVIGA		
	FSGGLGYPAVBAALGFMLPFLNFSFORKSLCVTAALVAGALAGVTLFSIP		
	VAILAGIVCGCLTALIQAFWQGAPEL"		
ORIGIN			
Alignment Scores:	5,54e-107	Length: 738	
Pred. No.:	1248.00	Matches: 245	
Score:	100.0%	Conservative: 0	
Percent Similarity:	100.0%	Mismatches: 0	
Best Local Similarity:	100.0%	Indels: 0	
Query Match:	6	Gaps: 0	
DB:			
US-10-073-293A-4 (1-245) x AX534668 (1-738)			
Qy	1	MetGluSerProThrProGlnProAlaProGlySerAlaThrPheMetGluGlyCysLys	20
Db	1	ATGGAAGCCCTACTCCACAGCCTGCTCTGGTTCGGCGACCTTCATGGAGGATGCAAA	60
Qy	21	AspSerLeuProIleValIleSerTyrIleProValAlaPheAlaPheGlyLeuAsnAla	40
Db	61	GACAGTTTACCGATTGTTATTAGTTATATTCGGTGGCCTTTGGCGTTCGATCGC	120
Qy	41	ThrArgLeuGlyPheSerProLeuGlnSerValPhePheSerCysIleIleTyrAlaGly	60
Db	121	ACCGCTCGGGATTCTCTCCCTCGAAAGCGTTTTTTCTCGCATCATTTATGCGAGGC	180
Qy	61	AlaSerGlnPheValIleThrAlaMetLeuAlaAlaGlySerSerLeuTyrIleAlaAla	80
Db	181	GGAGCCAGTTTCGTCATTACCGCATGCTGGCAGCGGAGTAGTTTGTGATTGCTGCA	240
Qy	81	LeuThrValMetAlaMetAspValArgHisValLeuTyrGlyProSerLeuArgSerArg	100
Db	241	CTGACCGTCATGGCAATGGATGTTCCCATGTTGTATGGCCCGTCACTCGTAGCCGT	300
Qy	101	IleIleGlnArgLeuGlnLysSerIlyThrAlaLeuTyrAlaPheGlyLeuThrAspGlu	120
Db	301	ATTATTACGCGTCTGCAAAATCGAAACCCGCCCTGTGGGCGTTTGGCGTACCGATGAG	360
Qy	121	ValPheAlaAlaAlaThrAlaLysLeuValArgAsnAsnArgTyrPserGluAsnTyr	140
Db	361	GTTTTTCGGCGCAACCGCAAACTGGTAGCAATAATCCCGCTGGAGCGAGACTGG	420
Qy	141	MetIleGlyIleAlaPheSerSerTyrPserTyrValPheGlyThrValIleGlyAla	160
Db	421	ATGATCGGCATTCGCTTCAGTTTCATGGTCATCGGGTATTGGTACGGTAAATAGGGCA	480
Qy	161	PheSerGlySerGlyLeuLeuGlnGlyTyrProAlaValGluAlaAlaLeuGlyPheMet	180
Db	481	TTCTCCGGCAGCGGCTTCGTCAGAGGTATCCCGCGGTTGAAGCTGCAATTAGGTTTATG	540
Qy	181	LeuProAlaLeuPheMetSerPheLeuLeuAlaSerPheGlnArgLysGlnSerLeuCys	200
Db	541	CTTCGGGACCTCTTATGAGTTTCTGCTCGCCTCTTCGAGCGCAACATCTCTTTCG	600
Qy	201	ValThrAlaAlaLeuValGlyAlaLeuAlaGlyValThrLeuPheSerIleProValala	220
Db	601	GTTACCGCAGCGGTAGTTGGTGCCCTTCGAGCGGTAAACGCTATTTTCTATTCCCGTCGCC	660
Qy	221	IleLeuAlaGlyIleValCysGlyCysLeuThrAlaLeuIleGlnAlaPheTyrGlnGly	240
Db	661	ATTTCTGGCAGGCATTGTTCTGTGGCTGCTCACTCGCTTAATCCAGGCATTCTGCAAGGA	720
Qy	241	AlaProAspGluLeu 245	
Db	721	GCGCCCGATGAGCTA 735	
RESULT 5			
D90891			
LOCUS	D90891 19150 bp DNA linear BCT 29-MAY-1997		
DEFINITION	E.coli genomic DNA, Kohara clone #445(60.2-60.6 min.).		
ACCESSION	D90891 AB001340		
VERSION	D90891.1 GI:18000054		
KEYWORDS	Complete and shotgun sequencing; HLYU; MG230; NRDE; ZK632.10; emrA; emrB; emrK; gshA; gshI; hnsB; mprA; nrdF; prou; prov; proW; proX; stpA; ygaC; ygaG; ygaH; yqjD; yzzM.		
SOURCE	Escherichia coli		
ORGANISM	Escherichia coli		
REFERENCE	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.		
AUTHORS	1 (sites)		
	Yamamoto,Y., Aiba,H., Baba,T., Hayashi,K., Inada,T., Isono,K., Itoh,T., Kimura,S., Kitagawa,M., Makino,K., Miki,T., Mitsuhashi,N., Mizobuchi,K., Mori,H., Nakade,S., Nakamura,Y., Nashimoto,H., Oshima,T., Oyama,S., Saito,N., Sampei,G., Satoh,Y., Sivasaundaram,S., Tagami,H., Takahashi,H., Takeda,J., Takemoto,K., Uehara,K., Wada,C., Yamagata,S. and Horiuchi,T.		
TITLE	Construction of a contiguous 874-kb sequence of the Escherichia coli -K12 genome corresponding to 50.0-68.8 min on the linkage map and analysis of its sequence features		
JOURNAL	DNA Res. 4 (2), 91-113 (1997)		
PUBMED	9205837		
REFERENCE	2 (sites)		
AUTHORS	Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C., Yamamoto,Y. and Yano,M.		
TITLE	The systematic sequencing of the Escherichia coli genome in Japan		
JOURNAL	Unpublished		
AUTHORS	3 (bases 1 to 19150)		
	Mori,H.		
TITLE	Direct Submission		
JOURNAL	Submitted (23-JAN-1997) Hirotada Mori, NARA Institute of Science and Technology, Res. & Edu. Center for Genetic Info.; 8916-5 Takayama, Ikoma, Nara 630-01, Japan		
COMMENT	(E-mail:hmori@etc.aist-nara.ac.jp, Tel:81-7437-2-5660, Fax:81-7437-2-5669) Collaboration Information: Project: The Japan E.coli genome DNA sequencing project Group: The Japan E.coli genome DNA sequencing group Members: (1995.4 - 1996.3) Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C., Yamamoto,Y. and Yano,M.		
	Headed by:		
	Name: Takashi Horiuchi		
	Address: National Institute of Basic Biology, Okazaki, 444, Japan		
	E-mail: kishori@nibb.ac.jp		
	Information operator:		
	Name: Hirotada Mori		
	Address: NARA Institute of Science and Technology, Ikoma, 630-01, Japan		

E-mail: hmori@gtc.aist-nara.ac.jp
 URL:
 The Japan E. coli genome database
 http:bsw3.aist-nara.ac.jp.
 Location/Qualifiers

FEATURES

source

1..19150
 /organism="Escherichia coli"
 /mol_type="genomic DNA"
 /strain="K12"
 /db_xref="taxon:562"
 /map="60.2 min"
 /clone="Kohara clone #445"
 /clone_lib="Kohara lambda miniset library"
 /note="Nucleotide position 2791307-2810456 from the
 initiation site of ThrA (0 min.).
 This clone is from Kohara lambda miniset library"
 complement(1..253)
 /gene="yzzM"
 complement(<1..>253)
 /gene="yzzM"
 /note="similar to [SwissProt Accession Number P39169]
 start codon is not identified yet"
 /codon_start=1
 /transl_table=11
 /product="UNKNOWN PROTEIN FROM 2D-PAGE (SPOT LM6)
 (FRAGMENT)."
 /protein_id="BAA16532.1"
 /db_xref="GI:1800055"
 /translation="GLFNFVKDAGEKLWDVATGQHKDDQAKVQEHNLKGTIPDADK
 VNIQADKATVVGDLGSQAQAKILVAVNGISIASVDD"
 complement(340..498)
 /gene="ZK632.10"
 complement(340..498)
 /gene="ZK632.10"
 /note="similar to [SwissProt Accession Number P34655]"
 /codon_start=1
 /transl_table=11
 /protein_id="BAA16533.1"
 /db_xref="GI:1800056"
 /translation="MGFWRVITIIPLPLGLLKGKGFMAFIINILLTLGLYIPGLIH
 AFWVQTRD"
 681..980
 /gene="HLYU"
 681..980
 /gene="HLYU"
 /note="similar to [SwissProt Accession Number P52695]"
 /codon_start=1
 /transl_table=11
 /product="TRANSCRIPTIONAL ACTIVATOR HLYU."
 /protein_id="BAA16534.1"
 /db_xref="GI:1800057"
 /translation="MTELAQLQASAEQAAALLKAMSHPKELLILCMLSGSPGTSAGEL
 TRTGLSASATSOHLARMDEGLDSQDQRIILYSIKNEAVNAIATLKNVYCP"
 complement(1561..1965)
 /gene="hnsB"
 complement(1561..1965)
 /gene="hnsB"
 /note="similar to [SwissProt Accession Number P30017]"
 /codon_start=1
 /transl_table=11
 /product="DNA-BINDING PROTEIN STPA."
 /protein_id="BAA16535.1"
 /db_xref="GI:1800058"
 /translation="MSVMQLSNIRTLRAMAREFSIDVLEEMLEKFRVVTKRRREE
 EQOQRELAERQEKISTVLEMLKADGINPELLGNSSAAAPRAGKKRQPRPAKYKFTDV
 NGTKYTWGGQRTPKPTAQLAEGKSLDDFLI"
 complement(3120..3476)
 /gene="ygaC"
 complement(3120..3476)
 /gene="ygaC"
 /note="similar to [SwissProt Accession Number P36931]
 start codon is not identified yet"

gene

CDS

/codon_start=1
 /transl_table=11
 /protein_id="BAA16536.1"
 /db_xref="GI:1800059"
 /translation="RGSNNYLRPDEVARVLEKVGFTVVDVYTKAYGYRRCNGYVYVNR
 EARMGRTALVIHETLKERSSTLAEPASDIKCDHYQQFPFLYLAGERHERHYGIPHGFS
 RVALERYLNGLFGEAS"
 3604..3945
 /gene="yqjD"
 3604..3945
 /gene="yqjD"
 /note="similar to [SwissProt Accession Number P42617]"
 /codon_start=1
 /transl_table=11
 /protein_id="BAA16537.1"
 /db_xref="GI:1800060"
 /translation="MGDHFNPRNRNDVDGVQDIQNDVNQLADSLSVLKSWSGSDAK
 GEAAARSAQALLKETRAMHGRTRVQQAARDVAGCADSFVRERPMCSVGTAAAVGI
 FIGALLSMRKS"
 4193..4846
 /gene="MG230"
 4193..4846
 /gene="MG230"
 /note="similar to [SwissProt Accession Number P47472]"
 /codon_start=1
 /transl_table=11
 /protein_id="BAA16538.1"
 /db_xref="GI:1800061"
 /translation="MRITTYTRNDVCQCHATKRAMENRGDFPEMINVDVRYPERAEALR
 AQGFQPLVVIAGDLWSGFRPDMINRLASSATRGQMSQLVYFSSSENTQRTFIERL
 GLPAVRIPLNERIRIQVDEPFIILVPSYGGGTAGVPRQVIRFLNDEHNRALLRGVI
 ASGNRNFGEAYGRAGDVIARKCGVPLYRFELMQTQSDIENVRKGVTEFWQRPQNA"
 4858..6414
 /gene="NRDE"
 4858..6414
 /gene="NRDE"
 /note="similar to [SwissProt Accession Number Q08698]"
 /codon_start=1
 /transl_table=11
 /product="RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE 2 ALPHA
 CHAIN (EC 1.17.4.1) (RIBONUCLEOTIDE REDUCTASE 2)."
 /protein_id="BAA16539.1"
 /db_xref="GI:1800062"
 /translation="MDYHALNMLNLYDSAGRIQFDKDRQADAFIATHVRPNVSFTS
 SOQRLNMLVNEGYYDESRYDRFVITLFTHTSGRFFOTFLGAWFYTSYTLK
 TDGKYLDFADRVMTVALTLAQGETLALQLTDEMLSGRFQFATPELNCCKQQRG
 ELVSCFLRIEDNMESIGRAVNSALQISKRGGVAFLLNLRAGAPIKKIENQSSGV
 IPVMKLEDAFSYANQLQAGARQAGAVILHAHPDILRFLDTKRENADEKIRIKTSLUG
 WIPDITFLAKENQAQALFSPYDVERVYKPFADVAISQHYDELVADERIRKYLNA
 RQFFORLARIQFESGYPIYVEDTVNRANPIAGRIINMSNLCSILQVNSASEYDENLD
 YTRTGHDISCNLSGLNIATMDSPPDFARTVETAVRGLTAVSDMSHRSVPSIAGNAA
 SHAIGLQGNLHGYLAREGLIAYGSPALDFTNLYFYAITWHALRTSMLARERGETTA
 GFKQSRYAQ"
 6272..6964
 /note="similar to [PIR Accession Number S34271]"
 /codon_start=1
 /transl_table=11
 /product="ribonucleoside-diphosphate reductase (EC
 1.17.4.1)"
 /protein_id="BAA16540.1"
 /db_xref="GI:1800063"
 /translation="MVRRKHVSPISISMPSQGMHCVPRCCWHANAVKPSPGSNHAM
 PSGEYFQVYQGNWQPKTAKVGLSTRSGITLPLTRMWAQLRDDVMRYGYLQYNLOAV
 PPTGSIYSINHATSSIHPIVAKVEIRKEKGTGVYYPAPMTNENLALYQDAYEIGE

gene

CDS

/codon_start=1
 /transl_table=11
 /protein_id="BAA16533.1"
 /db_xref="GI:1800056"
 /translation="MGFWRVITIIPLPLGLLKGKGFMAFIINILLTLGLYIPGLIH
 AFWVQTRD"
 681..980
 /gene="HLYU"
 681..980
 /gene="HLYU"
 /note="similar to [SwissProt Accession Number P52695]"
 /codon_start=1
 /transl_table=11
 /product="TRANSCRIPTIONAL ACTIVATOR HLYU."
 /protein_id="BAA16534.1"
 /db_xref="GI:1800057"
 /translation="MTELAQLQASAEQAAALLKAMSHPKELLILCMLSGSPGTSAGEL
 TRTGLSASATSOHLARMDEGLDSQDQRIILYSIKNEAVNAIATLKNVYCP"
 complement(1561..1965)
 /gene="hnsB"
 complement(1561..1965)
 /gene="hnsB"
 /note="similar to [SwissProt Accession Number P30017]"
 /codon_start=1
 /transl_table=11
 /product="DNA-BINDING PROTEIN STPA."
 /protein_id="BAA16535.1"
 /db_xref="GI:1800058"
 /translation="MSVMQLSNIRTLRAMAREFSIDVLEEMLEKFRVVTKRRREE
 EQOQRELAERQEKISTVLEMLKADGINPELLGNSSAAAPRAGKKRQPRPAKYKFTDV
 NGTKYTWGGQRTPKPTAQLAEGKSLDDFLI"
 complement(3120..3476)
 /gene="ygaC"
 complement(3120..3476)
 /gene="ygaC"
 /note="similar to [SwissProt Accession Number P36931]
 start codon is not identified yet"

gene

CDS

/codon_start=1
 /transl_table=11
 /protein_id="BAA16533.1"
 /db_xref="GI:1800056"
 /translation="MGFWRVITIIPLPLGLLKGKGFMAFIINILLTLGLYIPGLIH
 AFWVQTRD"
 681..980
 /gene="HLYU"
 681..980
 /gene="HLYU"
 /note="similar to [SwissProt Accession Number P52695]"
 /codon_start=1
 /transl_table=11
 /product="TRANSCRIPTIONAL ACTIVATOR HLYU."
 /protein_id="BAA16534.1"
 /db_xref="GI:1800057"
 /translation="MTELAQLQASAEQAAALLKAMSHPKELLILCMLSGSPGTSAGEL
 TRTGLSASATSOHLARMDEGLDSQDQRIILYSIKNEAVNAIATLKNVYCP"
 complement(1561..1965)
 /gene="hnsB"
 complement(1561..1965)
 /gene="hnsB"
 /note="similar to [SwissProt Accession Number P30017]"
 /codon_start=1
 /transl_table=11
 /product="DNA-BINDING PROTEIN STPA."
 /protein_id="BAA16535.1"
 /db_xref="GI:1800058"
 /translation="MSVMQLSNIRTLRAMAREFSIDVLEEMLEKFRVVTKRRREE
 EQOQRELAERQEKISTVLEMLKADGINPELLGNSSAAAPRAGKKRQPRPAKYKFTDV
 NGTKYTWGGQRTPKPTAQLAEGKSLDDFLI"
 complement(3120..3476)
 /gene="ygaC"
 complement(3120..3476)
 /gene="ygaC"
 /note="similar to [SwissProt Accession Number P36931]
 start codon is not identified yet"

gene

CDS

/codon_start=1
 /transl_table=11
 /protein_id="BAA16533.1"
 /db_xref="GI:1800056"
 /translation="MGFWRVITIIPLPLGLLKGKGFMAFIINILLTLGLYIPGLIH
 AFWVQTRD"
 681..980
 /gene="HLYU"
 681..980
 /gene="HLYU"
 /note="similar to [SwissProt Accession Number P52695]"
 /codon_start=1
 /transl_table=11
 /product="TRANSCRIPTIONAL ACTIVATOR HLYU."
 /protein_id="BAA16534.1"
 /db_xref="GI:1800057"
 /translation="MTELAQLQASAEQAAALLKAMSHPKELLILCMLSGSPGTSAGEL
 TRTGLSASATSOHLARMDEGLDSQDQRIILYSIKNEAVNAIATLKNVYCP"
 complement(1561..1965)
 /gene="hnsB"
 complement(1561..1965)
 /gene="hnsB"
 /note="similar to [SwissProt Accession Number P30017]"
 /codon_start=1
 /transl_table=11
 /product="DNA-BINDING PROTEIN STPA."
 /protein_id="BAA16535.1"
 /db_xref="GI:1800058"
 /translation="MSVMQLSNIRTLRAMAREFSIDVLEEMLEKFRVVTKRRREE
 EQOQRELAERQEKISTVLEMLKADGINPELLGNSSAAAPRAGKKRQPRPAKYKFTDV
 NGTKYTWGGQRTPKPTAQLAEGKSLDDFLI"
 complement(3120..3476)
 /gene="ygaC"
 complement(3120..3476)
 /gene="ygaC"
 /note="similar to [SwissProt Accession Number P36931]
 start codon is not identified yet"

Alignment Scores:

Pred. No.:	3,25e-105	Length:	19150
Score:	1248.00	Matches:	245
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	100.0%	Indels:	0
DB:	1	Gaps:	0

US-10-073-293A-4 (1-245) x D90891 (1-19150)

```
Qy      1 MetGluSerProThrProGlnProAlaProGlySerAlaThrPheMetGluGlyCysLys 20
Db      13090 ATGGAAGCCCTACTCCACAGCCTCTCTGGTTCGGCGACCTTCATGGAAGGATGCAAA 13149
Qy      21 AspSerLeuProIleValIleSerTyrIleProValAlaPheAlaPheGlyLeuAsnAla 40
Db      13150 GACAGTTTACCGATGTTATTAGTTATATTCGGTGGCCTTTGGCTCGGTCTGAATGCG 13209
Qy      41 ThrArgLeuGlyPheSerProLeuGluSerValPhePheSerCysIleIleTyrAlaGly 60
Db      13210 ACCCGCTCGGATTCCTCTCTCGAAAGCGTTTTTTTCTCTGTCATCATTTATGCGAGC 13269
Qy      61 AlaSerGlnPheValIleThrAlaMetLeuAlaGlySerSerLeuTrpIleAlaAla 80
Db      13270 GCGAGCCAGTTTCGTCATTACCGCGATGCTGCAGCCGGAGTAGTTTGTGGATTGCTGCA 13329
Qy      81 LeuThrValMetAlaMetAspValArgHisValLeuTyrGlyProSerLeuArgSerArg 100
Db      13330 CTGACCGTCATGGCAATGGATGTTCCCATGTGTGTATGGCCCGTCACTCGGTAGCCGT 13389
Qy      101 IleIleGlnArgLeuGlnLysSerIleThrAlaLeuTrpAlaPheGlyLeuThrAspGlu 120
Db      13390 ATTATTCAGCGCTCTGCAAAAATCGAAAAACCGCCCTGTGGCGTTTGGCCTGACGGATGAG 13449
Qy      121 ValPheAlaAlaAlaThrAlaLysLeuValArgAsnAsnArgArgTrpSerGluAsnTrp 140
Db      13450 GTTTTTCGGCGCGCAACCGCAAACTGGTACGCAATAATCCGCCGTGGAGCGAGAACTGG 13509
Qy      141 MetIleGlyIleAlaPheSerSerTrpSerTrpValPheGlyThrValIleGlyAla 160
Db      13510 ATGATCGGCATTGCTTCAGTTTCATGTCATCGTGGGTATTTGGTACGGTAAATAGGGCA 13569
Qy      161 PheSerGlySerGlyLeuLeuGlnGlyTyrProAlaValGluAlaAlaLeuGlyPheMet 180
Db      13570 TTTCTCCGCGCAGCGGCTTGCTCGCAAGTTATCCCGCGGTTCGAAGCTGCAATTAGGTTTATG 13629
Qy      181 LeuProAlaLeuPheMetSerPheLeuLeuAlaSerPheGlnArgLysGlnSerLeuCys 200
Db      13630 CTTCGGGCACATCTTTATAGATTTCCTGCTCGCCTCTTTCCAGCGCAACAATCTCTTTGC 13689
Qy      201 ValThrAlaAlaLeuValGlyAlaLeuAlaGlyValThrLeuPheSerIleProValAla 220
Db      13690 GTTACCGCAGGTTAGTTGGTGCCCTTCGACGGCGTAACGCTATTTCTATTCCCGTCGCC 13749
Qy      221 IleLeuAlaGlyIleValCysGlyCysLeuThrAlaLeuIleGlnAlaPheTrpGlnGly 240
Db      13750 ATTCTGGCAGGCATTGTCTGTGGCTGCTCACTGCGTTAATCCAGGCATTCTGGCAAGGA 13809
Qy      241 AlaProAspGluLeu 245
Db      13810 GCGCCCGATGAGCTA 13824
```

RESULT 6

U00096_27

WPCOMMENT

Sequence split into 47 fragments LOCUS U00096 Accession U00096

Fragment Name	Begin	End
U00096_00	1	110000
U00096_01	100001	210000
U00096_02	200001	310000
U00096_03	300001	410000
U00096_04	400001	510000
U00096_05	500001	610000
U00096_06	600001	710000
U00096_07	700001	810000
U00096_08	800001	910000
U00096_09	900001	1010000
U00096_10	1000001	1110000
U00096_11	1100001	1210000
U00096_12	1200001	1310000
U00096_13	1300001	1410000

```
U00096_14 1400001 1510000
U00096_15 1500001 1610000
U00096_16 1600001 1710000
U00096_17 1700001 1810000
U00096_18 1800001 1910000
U00096_19 1900001 2010000
U00096_20 2000001 2110000
U00096_21 2100001 2210000
U00096_22 2200001 2310000
U00096_23 2300001 2410000
U00096_24 2400001 2510000
U00096_25 2500001 2610000
U00096_26 2600001 2710000
U00096_27 2700001 2810000
U00096_28 2800001 2910000
U00096_29 2900001 3010000
U00096_30 3000001 3110000
U00096_31 3100001 3210000
U00096_32 3200001 3310000
U00096_33 3300001 3410000
U00096_34 3400001 3510000
U00096_35 3500001 3610000
U00096_36 3600001 3710000
U00096_37 3700001 3810000
U00096_38 3800001 3910000
U00096_39 3900001 4010000
U00096_40 4000001 4110000
U00096_41 4100001 4210000
U00096_42 4200001 4310000
U00096_43 4300001 4410000
U00096_44 4400001 4510000
U00096_45 4500001 4610000
U00096_46 4600001 4639675

Continuation (28 of 47) of U00096 from base 2700001 (U00096 Escherichia coli K-12 MG1655)

Alignment Scores:
Pred. No.: 2,89e-104 Length: 110000
Score: 1248.00 Matches: 245
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 1 Gaps: 0

US-10-073-293A-4 (1-245) x U00096_27 (1-110000)

Qy      1 MetGluSerProThrProGlnProAlaProGlySerAlaThrPheMetGluGlyCysLys 20
Db      107639 ATGGAAGCCCTACTCCACAGCCTGCTCTGGTTCGGCGACCTTCATGGAAGGATGCAAA 107698
Qy      21 AspSerLeuProIleValIleSerTyrIleProValAlaPheAlaPheGlyLeuAsnAla 40
Db      107699 GACAGTTTACCGATTGTTATTAGTTATATTCGGTGGCCTTTGGCTCGGTCTGAATGCG 107758
Qy      41 ThrArgLeuGlyPheSerProLeuGluSerValPhePheSerCysIleIleTyrAlaGly 60
Db      107759 ACCCGCTCGGATTCCTCTCTCGAAAGCGTTTTTTTCTCTGTCATCATTTATGCGAGC 107818
Qy      61 AlaSerGlnPheValIleThrAlaMetLeuAlaGlySerSerLeuTrpIleAlaAla 80
Db      107819 GCGAGCCAGTTTCGTCATTACCGCGATGCTGGCAGCCGGAGTAGTTTGTGGATTGCTGCA 107878
Qy      81 LeuThrValMetAlaMetAspValArgHisValLeuTyrGlyProSerLeuArgSerArg 100
Db      107879 CTGACCGTCATGGCAATGGATGTTGCGCATGTGTGTATGGCCCGTCACTCGGTAGCCGT 107938
Qy      101 IleIleGlnArgLeuGlnLysSerIleThrAlaLeuTrpAlaPheGlyLeuThrAspGlu 120
Db      107939 ATTATTCAGCGCTCTGCAAAAATCGAAAAACCGCCCTGTGGCGGTTTGGCCTGACGGATGAG 107998
Qy      121 ValPheAlaAlaAlaThrAlaLysLeuValArgAsnAsnArgArgTrpSerGluAsnTrp 140
Db      107999 GTTTTTCGCGCGCAACCGCAAAAATCGTACGCAATAATCGCCCGTGGAGCGAGAACTGG 108058
```

Qy 141 MetIleGlyIleAlaPheSerSerTrpSerTrpValPheGlyThrValIleGlyAla 160
Db 108059 ATGATCGGCATTGCTTCAGTTTCATCGTGGGTATTTGGTACGGTAATAGGGCA 108118
Qy 161 PheSerGlySerGlyLeuLeuGlnGlyTyrProAlaValGluAlaLeuGlyPheMet 180
Db 108119 TTCTCCGGCAGCGCTTCTCTGCAAGGTATCCCGCGTTGAAGCTGCATTAGGTTTATG 108178
Qy 181 LeuProAlaLeuPheMetSerPheLeuLeuAlaSerPheGlnArgLysGlnSerLeuCys 200
Db 108179 CTTCGGGCACCTCTTATGAGTTTCTGCTCGCCTCTTTCCAGCGCAACAACTCTTTGC 108238
Qy 201 ValThrAlaAlaLeuValGlyAlaLeuAlaGlyValThrLeuPheSerIleProValAla 220
Db 108239 GTTACCGCAGCGTTAGTTGGTGGCCCTTGCAGGCGTAACGCTATTTTCTATTCCCGTCGCC 108298
Qy 221 IleLeuAlaGlyIleValCysGlyCysLeuThrAlaLeuIleGlnAlaPheTrpGlnGly 240
Db 108299 ATTCTGGCAGGCATTGTCTGTGGCTCACTGCGTTATCCAGGCAATCTTGGCAAGGA 108358
Qy 241 AlaProAspGluLeu 245
Db 108359 GCGCCCGATGAGCTA 108373

RESULT 7

U00096_28

WPCOMMENT

Sequence split into 47 fragments LOCUS U00096 Accession U00096

Fragment Name	Begin	End
U00096_00	1	110000
U00096_01	100001	210000
U00096_02	200001	310000
U00096_03	300001	410000
U00096_04	400001	510000
U00096_05	500001	610000
U00096_06	600001	710000
U00096_07	700001	810000
U00096_08	800001	910000
U00096_09	900001	1010000
U00096_10	1000001	1110000
U00096_11	1100001	1210000
U00096_12	1200001	1310000
U00096_13	1300001	1410000
U00096_14	1400001	1510000
U00096_15	1500001	1610000
U00096_16	1600001	1710000
U00096_17	1700001	1810000
U00096_18	1800001	1910000
U00096_19	1900001	2010000
U00096_20	2000001	2110000
U00096_21	2100001	2210000
U00096_22	2200001	2310000
U00096_23	2300001	2410000
U00096_24	2400001	2510000
U00096_25	2500001	2610000
U00096_26	2600001	2710000
U00096_27	2700001	2810000
U00096_28	2800001	2910000
U00096_29	2900001	3010000
U00096_30	3000001	3110000
U00096_31	3100001	3210000
U00096_32	3200001	3310000
U00096_33	3300001	3410000
U00096_34	3400001	3510000
U00096_35	3500001	3610000
U00096_36	3600001	3710000
U00096_37	3700001	3810000
U00096_38	3800001	3910000
U00096_39	3900001	4010000
U00096_40	4000001	4110000
U00096_41	4100001	4210000
U00096_42	4200001	4310000
U00096_43	4300001	4410000

U00096_44 4400001 4510000
U00096_45 4500001 4610000
U00096_46 4600001 4639675

Continuation (29 of 47) of U00096 from base 2800001 (U00096 Escherichia coli K-12 MG1615)

Alignment Scores:
Pred. No.: 2.89e-104 Length: 110000
Score: 1248.00 Matches: 245
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 1 Gaps: 0

US-10-073-293A-4 (1-245) x U00096_28 (1-110000)

Qy 1 MetGluSerProThrProGlnProAlaProGlySerAlaThrPheMetGluGlyCysAla 20
Db 7639 ATGGAAAGCCCTACTCCACAGCCCTGCTCTGGTTCGGCGACCTTCATGGAGAGTAGCMAA 7698
Qy 21 AspSerLeuProIleValIleSerTyrIleProValAlaPheAlaPheGlyLeuAsnAla 40
Db 7699 GACAGTTTACCGATTGTTATTAGTTATATTCGGTGGCCCTTGGGTTCCGATCATCATTTATG 7758
Qy 41 ThrArgLeuGlyPheSerProLeuGluSerValPhePheSerCysIleIleTyrAlaGly 60
Db 7759 ACCCGTCTGGGATTCTCTCTCGAAAGCGTTTTTTTCTCTCGATCATCATTTATGACAGC 7818
Qy 61 AlaSerGlnPheValIleThrAlaMetLeuAlaGlySerSerLeuTrpIleAlaAla 80
Db 7819 GCGAGCCAGTTCGTCATTACCGGATGCTGGAGCCGGAGTAGTTGTGGATTGCTGCA 7878
Qy 81 LeuThrValMetAlaMetAspValArgHisValLeuTyrGlyProSerLeuArgSerArg 100
Db 7879 CTGACCGTCATGCAATGGATGTTCCGCATGTGTGTATGTCGCCGTCTCCTGCTAGCGT 7938
Qy 101 IleIleGlnArgLeuGlnLysSerLysThrAlaLeuTrpAlaPheGlyLeuThrAspGlu 120
Db 7939 ATTATTTCAGCGTCTGCAGAAATCGAAACCGCCCTGTGGCGGTTTGGCTGACGAGATGAG 7998
Qy 121 ValPheAlaAlaAlaThrAlaLysLeuValArgAsnArgArgTrpSerGluAsnTrp 140
Db 7999 GTTTTTTCCCGCCGACCCGCAAACTGTACGCAATATCGCCCTGGAGCGAGACTGG 8058
Qy 141 MetIleGlyIleAlaPheSerSerTrpSerTrpValPheGlyThrValIleGlyAla 160
Db 8059 ATGATCGGCATTGCCITTCAGTTTCATGTCATCGTGGGTATTTGGTACGGTAATAGGGCA 8118
Qy 161 PheSerGlySerGlyLeuLeuGlnGlyTyrProAlaValGluAlaLeuGlyPheMet 180
Db 8119 TTCTCCGGCAGCGCTTGTCTGCAAGGTTATCCCGCCGCTTGAAGCTGCATTAGGTTTATG 8178
Qy 181 LeuProAlaLeuPheMetSerPheLeuLeuAlaSerPheGlnArgLysGlnSerLeuCys 200
Db 8179 CTTCCGGCAGCTCTTTATGAGTTTCTGCTCGCCTCTTTCCAGCGCAACAACTCTCTTTCG 8238
Qy 201 ValThrAlaAlaLeuValGlyAlaLeuAlaGlyValThrLeuPheSerIleProValAla 220
Db 8239 GTTACCGCAGCGTTAGTTGGTGGCCCTTGCAGGCGTAACGCTATTTTCTATTCCGTCGCC 8298
Qy 221 IleLeuAlaGlyIleValCysGlyCysLeuThrAlaLeuIleGlnAlaPheTrpGlnGly 240
Db 8299 ATTCTGGCAGGCATTGTCTGTGGCTGCTCACTGCGTTATCCAGGCAATCTTGGCAAGGA 8358
Qy 241 AlaProAspGluLeu 245
Db 8359 GCGCCCGATGAGCTA 8373

RESULT 8

AE005174_35

WPCOMMENT

Sequence split into 56 fragments LOCUS AE005174 Accession AE005174
Fragment Name Begin End
AE005174_00 1 110000

AE005174_01	100001	210000	
AE005174_02	200001	310000	
AE005174_03	300001	410000	
AE005174_04	400001	510000	
AE005174_05	500001	610000	
AE005174_06	600001	710000	
AE005174_07	700001	810000	
AE005174_08	800001	910000	
AE005174_09	900001	1010000	
AE005174_10	1000001	1110000	
AE005174_11	1100001	1210000	
AE005174_12	1200001	1310000	
AE005174_13	1300001	1410000	
AE005174_14	1400001	1510000	
AE005174_15	1500001	1610000	
AE005174_16	1600001	1710000	
AE005174_17	1700001	1810000	
AE005174_18	1800001	1910000	
AE005174_19	1900001	2010000	
AE005174_20	2000001	2110000	
AE005174_21	2100001	2210000	
AE005174_22	2200001	2310000	
AE005174_23	2300001	2410000	
AE005174_24	2400001	2510000	
AE005174_25	2500001	2610000	
AE005174_26	2600001	2710000	
AE005174_27	2700001	2810000	
AE005174_28	2800001	2910000	
AE005174_29	2900001	3010000	
AE005174_30	3000001	3110000	
AE005174_31	3100001	3210000	
AE005174_32	3200001	3310000	
AE005174_33	3300001	3410000	
AE005174_34	3400001	3510000	
AE005174_35	3500001	3610000	
AE005174_36	3600001	3710000	
AE005174_37	3700001	3810000	
AE005174_38	3800001	3910000	
AE005174_39	3900001	4010000	
AE005174_40	4000001	4110000	
AE005174_41	4100001	4210000	
AE005174_42	4200001	4310000	
AE005174_43	4300001	4410000	
AE005174_44	4400001	4510000	
AE005174_45	4500001	4610000	
AE005174_46	4600001	4710000	
AE005174_47	4700001	4810000	
AE005174_48	4800001	4910000	
AE005174_49	4900001	5010000	
AE005174_50	5000001	5110000	
AE005174_51	5100001	5210000	
AE005174_52	5200001	5310000	
AE005174_53	5300001	5410000	
AE005174_54	5400001	5510000	
AE005174_55	5500001	5528445	

Continuation (36 of 56) of AE005174 from base 3500001 (AE005174_35 (1-110000))

Alignment Scores:			
Pred. No.:	3,586-104	Length:	110000
Score:	1247.00	Matches:	244
Percent Similarity:	100.0%	Conservative:	1
Best Local Similarity:	99.6%	Mismatches:	0
Query Match:	99.9%	Indels:	0
DB:	1	Gaps:	0

US-10-073-293A-4 (1-245) x AE005174_35 (1-110000)

Qy	1	MetGluSerProThrProGlnProAlaThrPheMetGluGlyCysLys	20
Db	101065	ATGGAAGCCCTACTCCAGCCTCTCTGTTCCGCGACCTTCATGGAAGGATGCAAA	101124
Qy	21	AspSerLeuProIleValIleSerTyrIleProValAlaPheAlaPheGlyLeuAsnAla	40

Db	101125	GACAGTTTACCGATTGTTATTAGTTATATCCCGTGGCTTTTGGCTTCGATGCG	101184
Qy	41	ThrArgLeuGlyPheSerProLeuGluSerValPhePheSerCysIleIleTyrAlaGly	60
Db	101185	ACCGCTCTGGGATTCTCTCTCTCGAAAGCGTTTTTTTCTCTGCATCATTTATGCAAGC	101244
Qy	61	AlaSerGlnPheValIleThrAlaMetLeuAlaAlaGlySerSerLeuTrpIleAlaAla	80
Db	101245	GCGAGCCAGTTCGTCATTACCGCATGTTGTCACAGAGTAGTTTGTGGTTGCTGCA	101304
Qy	81	LeuThrValMetAlaMetAspValArgHisValLeuTyrGlyProSerLeuArgSerArg	100
Db	101305	CTGACCGTCATGGCAATGGATGTTTCGCATGTTTGTATGCGCCGTCACTCGCTAGCCGT	101364
Qy	101	IleIleGlnArgLeuGlnLysSerLysThrAlaLeuTrpAlaPheGlyLeuThrAspGlu	120
Db	101365	ATTATTCAGCGCTCTGCAAAATCGAAACCGCCCTATGGGGCTTTGGCCTACCGATGAG	101424
Qy	121	ValPheAlaAlaAlaThrAlaLysLeuValArgAsnAsnArgTgTgSerGluAsnTrp	140
Db	101425	GTTTTTGGCCGCCACCGCAAACTGTATGCGCAATATCCCGCTGGAGCGAAGACTGG	101484
Qy	141	MetIleGlyIleAlaPheSerSerTrpSerTrpValPheGlyThrValIleGlyAla	160
Db	101485	ATGATCGGCATTGCTTCAGTTTCATGTCGTTGGTATTTGTACGTTAATAGGGCA	101544
Qy	161	PheSerGlySerGlyLeuLeuGlnGlyTyrProAlaValGluAlaAlaLeuGlyPheMet	180
Db	101545	TTCTCCGCGACGGCTTGTCTGCAAGGTATATCCCGCTTGAAGCTGCATTAGTTTATG	101604
Qy	181	LeuProAlaLeuPheMetSerPheLeuLeuAlaSerPheGlnArgLysGlnSerLeuCys	200
Db	101605	CTTCCGCGACTCTTTATGAGTTTCTCTGCTCGCTCTTTTCAGCGCAACATCTCTTTC	101664
Qy	201	ValThrAlaAlaLeuValGlyAlaLeuAlaGlyValThrLeuPheSerIleProValAla	220
Db	101665	GTTACCGCAGCGTTAGTTGGTGCCCTTGCAGCGCTTACGCTATTTTCTATTCCGCTGCC	101724
Qy	221	IleLeuAlaGlyIleValCysGlyCysLeuThrAlaLeuIleGlnAlaPheTrpGlnGly	240
Db	101725	ATTCTGCGAGGCATTGTCTGTGGCTGCTCTACTGCTTAATCCAGGCATTTCTGCAAGGA	101784
Qy	241	AlaProAspGlnLeu 245	
Db	101785	GCGCCCGATGAGCTA 101799	

Sequence split into 56 fragments LOCUS AE005174 Accession AE005174

Fragment Name	Begin	End
AE005174_00	1	110000
AE005174_01	100001	210000
AE005174_02	200001	310000
AE005174_03	300001	410000
AE005174_04	400001	510000
AE005174_05	500001	610000
AE005174_06	600001	710000
AE005174_07	700001	810000
AE005174_08	800001	910000
AE005174_09	900001	1010000
AE005174_10	1000001	1110000
AE005174_11	1100001	1210000
AE005174_12	1200001	1310000
AE005174_13	1300001	1410000
AE005174_14	1400001	1510000
AE005174_15	1500001	1610000
AE005174_16	1600001	1710000
AE005174_17	1700001	1810000
AE005174_18	1800001	1910000
AE005174_19	1900001	2010000
AE005174_20	2000001	2110000
AE005174_21	2100001	2210000

RESULT 9
AE005174_36
WPCOMMENT


```

Db      34495  GCGCCCGATGAGCTA 34509
|||||
AE016765      305325 bp      DNA      linear      BCT 24-DEC-2002
LOCUS      Escherichia coli CFT073 section 11 of 18 of the complete genome.
ACCESSION      AE016765 AB014075
VERSION      AE016765.1 GI:26109395
KEYWORDS
SOURCE      Escherichia coli CFT073
ORGANISM      Escherichia coli CFT073
REFERENCE      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
AUTHORS      Enterobacteriaceae; Escherichia.
REFERENCE      1 (bases 1 to 305325)
AUTHORS      Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,
Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,
Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,
Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.
TITLE      Extensive Mosaic Structure Revealed by the Complete Genome Sequence
of Uropathogenic Escherichia coli
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 99 (26), 17020-17024 (2002)
PUBMED      12471157
REFERENCE      2 (bases 1 to 305325)
AUTHORS      Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,
Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,
Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,
Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.
TITLE      Direct Submission
JOURNAL      Submitted (20-JUN-2002) Genetics Laboratory, University of
Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
source
    1..305325
        /organism="Escherichia coli CFT073"
        /mol_type="genomic DNA"
        /strain="CFT073"
        /db_xref="taxon:199310"
        complement(145..438)
        /locus_tag="c3147"
        complement(145..438)
        /locus_tag="c3147"
        /codon_start=1
        /transl_table=11
        /product="Hypothetical protein"
        /protein_id="AA081599.1"
        /db_xref="GI:26109396"
        /translation="MDLTPFLHALCAVAQQLVCLFTGNWAYGAIACGTFPIAREHTQ
        AETRLMEFGHGRINPMWQFDPVRVMDVASLMDFAFVAVVACLLWLLVNRG"
        complement(481..1521)
        /locus_tag="c3148"
        complement(481..1521)
        /locus_tag="c3148"
        /notes="Residues 110 to 233 of 346 are 31.25 pct identical
        to residues 486 to 611 of 669 from GenPept.129 :
        >gb|AA03286.1| (AF063097) gph [Enterobacteria phage P2]"
        /codon_start=1
        /transl_table=11
        /product="Hypothetical protein"
        /protein_id="AA081600.1"
        /db_xref="GI:26109397"
        /translation="MATSTVIPPDIKTLKSDVSKLRNDQGSYATKLIVDSKDEIVGDN
        TSNICWQFLPTSGAIFGRKLRSFTRTAGVEDAYCELYLKKWIDPSGNAMRLNLDNG
        TAWCPQFLNYGTMIPGDSGLYKMGNCILMSYKRGSNALIKFDYDTDTLQIIVYNGA
        STWLTNQTGAYAGVTAQLMGNSRRPVVYEVGVDGGAYMFVAQKNTDITVYMLSVNGA
        HATAFNQHSRDLKDNILQVLDNATDIRKMGNYTKLKENGMPVAGVTAQDALPAIRP
        VVGSAMTKYQDQSGSGEQRITVTDYSGVTGLLVQVARESDDRITALEEENAEIAR
        RLSIAEAAALASK"
        complement(1531..1812)
        /locus_tag="c3149"
        complement(1531..1812)
        /locus_tag="c3149"
        /codon_start=1
        /transl_table=11
gene
CDS

```

```
/product="Hypothetical protein"
/protein_id="AAN81601.1"
/db_xref="GI:26109398"
/translacion="MKDLTKFADRADPFAFMESIGYVYDDESMOGDILLIDVIGNVYKE
TGEITDEGEPVCVKEDGYFVNVRILINDAKKSSIFDKYALVVEHQLRGM"
complement (1809..4238)
/locus_tag="c3150"
/locus_tag="c3150"
/locus_tag="c3150"
/locus_tag="c3150"
/notes="Residues 118 to 438 of 809 are 30.28 pct identical
to residues 291 to 638 of 973 from GenPept.129 :
>gb|AA056007.1|AE005333.4 (AE005333) putative membrane
protein of prophage CP-933X [Escherichia coli O157:H7
EDL933]"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAN81602.1"
/db_xref="GI:26109399"
/translacion="MPEDYGVDMAAVKISGVLGKAGKPIQNCITQLKAKRNTTV
LVNTVASENPDEAGRYNDVEYGYQYVILLVEGPPSHAGTIVYEGSRPGLNDPLG
AMTDDVMPALRFRFEMVBEAARNAEASQSAKAKKSETAAASAKNAKTSKTAAS
NSAQAATTSOTASENSATAKCKSETNAKSETAKTSETNAKSSQTAKTSETNAKAS
ETAANKSOVAASQESAASGATSATGSAANAQSAKTSKTAASQAKTSKTAASQAK
NAKASETAANKSOVAASQESAASGATSATGSAANAQSAKTSKTAASQAKTSKTAAS
ANSKASAAQTAAKASADEAREYASQAEPYKQVLPDVMIPFNDSILDMITGFP
SYKIVIGDDEITMPGDKVFKRASTAYINKGVFSVAKIDEPREKEGLIIEGOR
TNYPVKSNIPAEWTSNIDKNTNGVDFGFGSVAKMKTNDKMTGQSSALSLHTCSASR
GIDVSGDKNYCTSCRYKADGLRCLRPFKYDGSVYTFGLDAYLTFTGLIIEIKTGA
ANRIATAATDPVTFIYFATIEAVEGETLIGAMIQYAPKKGVTBAGDYIYIATPQ
FENGSGASSFYITTTVPATRSSDMVTIPTENNYNRELTLCLVEVRNMGDIPPNVAPR
IFDPSGVPYIESITATPTEKYYQYLMQTYKASTYTVSSVFAFGTDRVKETIGFN
IYSDTKRVVNGEATKMTKMTGVTKTRTPIRIGQATSGTRHLFGHLRLRLWHKE
LTDAMQGESIK"
/locus_tag="c3843"
/locus_tag="c3151"
/locus_tag="c3843"
/locus_tag="c3151"
/locus_tag="c3151"
/notes="Residues 15 to 146 of 146 are 45.32 pct identical
to residues 36 to 174 of 206 from SwissProt.40 :
>sp|P03774|Y206_LAMED_HYPOTHETICAL_PROTEIN_ORP206"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAN81603.1"
/db_xref="GI:26109400"
/translacion="MAFVSLVFAAVWLDPALVSLVFAAFWLLAAAVADPVAEAPAA
ALSLMAAATWFAAVSLALAFVSLVFAAVWLDPALVSLVFAAVSLVFAAVSLVFAAV
AEFSDAVCEVAACALLAFVSEVFAAFDDAAVSDFFAAAL"
complement (4222..4932)
/insertion_seq="IS1541A-like"
complement (4329..4838)
/locus_tag="c3152"
complement (4329..4838)
/locus_tag="c3152"
/locus_tag="c3152"
/notes="Residues 18 to 169 of 169 are 95.39 pct identical
to residues 1 to 152 of 152 from SwissProt.40 :
>sp|Q57334|T200_SALTY_Transposase_for_insertion_sequence
element_1S200"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAN81604.1"
/db_xref="GI:26109401"
/translacion="MRSGNCKCQTRNQKGVPMGNEKSLATRNWCKHIVPAPYRRQ
VFREKRAIGSLIKCKEWSKVIIEBCCADHIMLWVEIPPMKSVSGMGLYKGS
SLMLYFQFGLKFKYRNFRCRGYVYTVGRNTAKIQDYIKRHEEDKMGEOQLSIPY
PGSPFTGRK"
complement (4984..5583)
/locus_tag="c3153"
complement (4984..5583)
/locus_tag="c3153"
/function="putative membrane; Other or Phage or Prophage
```

```
Related"
/notes="Escherichia coli O157:H7 ortholog: z1917"
/codon_start=1
/transl_table=11
/product="Putative outer membrane protein of prophage"
/protein_id="AAN81605.1"
/db_xref="GI:26109402"
/translacion="MRKLCAVILSAVVLVAAGTPASAAEHQSTLSAGYLQTHDMPG
SDLLKGVNRYVEFTDLGLVTSFYANAKDEOKTHYSDTRMHEDSVNRNRPMSMAG
SVLVNEMFSAYAMAGMAYSRVSTSGDYLRVTDNCKTKHDVLTGSDDRHSNTSLAW
GAGVQNPSTESVALDLAYEGSGSDWRTDGFIVGVGYKP"
complement (5651..9130)
/locus_tag="c3154"
complement (5651..9130)
/locus_tag="c3154"
/locus_tag="c3154"
/function="putative structure; Structural component Phage
or Prophage Related"
/notes="Escherichia coli O157:H7 ortholog: z2145"
/codon_start=1
/transl_table=11
/product="Putative tail component of prophage"
/protein_id="AAN81606.1"
/db_xref="GI:26109403"
/translacion="MGKSGSKGHTPREAKDNLSQMSLVIDAISEGVEGVPDGLKS
VILNSTPLVDSQNTNIGSVTVVFRAGEQQTPEGSESGSETVLGTEVYDTPITR
TITSANIDRLRFTFGVALVETTSKGRNLSVRLLVQIQNGGWTVEKOTIKGKTT
SOYLASVVDNLPFRPNIRMRMTDSTTDQNKTLSSYTEIIDVKOCYPNTALV
GVQVDSQFQSQSVSRNYHLRGLIQLVPSNYPQTRQYSGIWDGTFPAPYNNMAWCL
WDMLTRYGMKRLGAAVDQKVALYVIGYCDQSPDGGFGGTSPRITCNALWTTQK
ADMLDFSCAMRCMPVNGQITLTFVQDRPSDKVTYNNRSNVNMPDGAAPRYSFSL
KDRHNAVENMIDPDNGWATELVDQTAIARYGRNVTKMDAFGCTSRGAHRAGLW
LITLLELTQVDFSVGAELRHVPQDIIEICDDYAGISTSGRLAVNSQTRTLTLD
REITLPSGGTLLISLVDGSGNPVSVEQSVTDGVKVSVPDGVSAVSWGLKLP
RQLRQVSIENDDGTATTAUVHUPKEALVDNGAHFDEGQSGTNGVTPPAVQHL
TAEVTADSGEYQLARWDTPKVGVSVFMLRLTVAADDSERLVTARTTETTRFRQ
LALGSYRLTVRAANAWGQGDPAVSFRYIAAPASRIELTPGYFQITATPHLAVYDP
TVQFEFVSEKRIADIQRTTARYLTALYIAASINIKPHDYYFYIRTSVTGKS
AFYEAQVQSDSDASGLDFFPKIGKSHLAQELWQIDNQLAPDLAIBRTSIDVSN
EITQVTKKLEQDQSAIQIQKQVQDNTNNLSMWAVKLOQMODGRLYIAGIGAIEN
TPDGMQSVQLAADRIAMVPANGTKPMFGQDQIPMNEVELKYLITAPITTSQGNP
PAFSLTPDGLRTAKWADISGNVANSGLNNVTINENCRVLGKLSAQIEGDLVKTVG
KAFPRDSRAPERPWSGTIVRIYDQDFDQIPIPAFSGAKHERHETDIYSSCLRI
VRKNGAEIYRNTALDNTLIYSGVIDMPAGHGMTELSFSASLWNNNYPTASISDLLV
VNMKATAGISIS"
complement (9191..9823)
/locus_tag="c3155"
complement (9191..9823)
/locus_tag="c3155"
/function="putative structure; Structural component Phage
or Prophage Related"
/notes="Escherichia coli O157:H7 ortholog: z0979"
/codon_start=1
/transl_table=11
/product="Putative tail component of prophage"
/protein_id="AAN81607.1"
/db_xref="GI:26109404"
/translacion="MARIQYGLDQRFGRIDLRVKTGAERIALATQLPAFROKLS
GWYQRIARQDVSTSGITAGLHETLPDGAIVHVPVRAKSGSGVFIIVGAAAIAGS
FFTAGATLAANGAAIAGGGGTILFSLGASVWLVGVAQMLAPKARTPTRTQTDNGKON
```

Alignment Scores:

Pred. No.:	1,28e-103	Length:	305325
Score:	1247.00	Matches:	244
Percent Similarity:	100.0%	Conservative:	1
Best Local Similarity:	99.6%	Mismatches:	0
Query Match:	99.9%	Indels:	0
DB:	1	Gaps:	0

US-10-073-293A-4 (1-245) x AE016765 (1-305325)

Qy 1 MetGluSerProThrProGlnProAlaProGlySerAlaThrPheMetGluGlyCysLeu 20

Db 68105 ATGGAAGGCCCTACTCCACAGCCCTGCTCTGTTCCGGCGACCTTCTGGAAGGATGCAAA 68164

Qy 21 AspSerLeuProIleValIleSerTyrIleProValAlaPheAlaPheGlyLeuAsnAla 40
Db 68165 GACAGTTTACCGATTGTATTAGTTATATTCGGTGGCCTTTGGCTTCGGTCTGAATGCG 68224
Qy 41 ThrArgLeuGlyPheSerProLeuGluSerValPhePheSerCysIleIleTyrAlaGly 60
Db 68225 ACCCGTCTGGGATTCTCTCCTCTCGAAGCGTTTTTCTCTCGATCAATTAATGACAGC 68284
Qy 61 AlaSerGlnPheValIleThrAlaMetLeuAlaAlaGlySerSerLeuTrpIleAlaAla 80
Db 68285 GCGAGCCAGTTTCGTCATTTACCGCATGCTGGCAGCCGGAGTAGTTTGGTGGTTGCTGCA 68344
Qy 81 LeuThrValMetAlaMetAspValArgHisValLeuTyrGlyProSerLeuArgSerArg 100
Db 68345 CTGACCGTCATGCGCAATGAGTGTTCGCCCATGTTTTGTATGGCCCGTCACTCGCTAGCCGT 68404
Qy 101 IleIleGlnArgLeuGlnLysSerLysThrAlaLeuTrpAlaPheGlyLeuThrAspGlu 120
Db 68405 ATTATTACGCTCTGCAAAATCGAAACCCCTGTGGGCGTTTGGCTGACCGATGAG 68464
Qy 121 ValPheAlaAlaAlaThrAlaLysLeuValArgAsnAsnArgTrpSerGluAsnTrp 140
Db 68465 GTTTTGGCCGCGCAACCGCAAACTGCTACGCAATTAATCGCCGCTGGAGCGAAACTGG 68524
Qy 141 MetIleGlyIleAlaPheSerSerTrpSerTrpValPheGlyThrValIleGlyAla 160
Db 68525 ATGATCGGCATTTGCCCTTCAGTTTCATGGTCAATCGTGGGTATTGTGTACGTTAATAGGGCA 68584
Qy 161 PheSerGlySerGlyLeuLeuGlnGlyTyrProAlaValGluAlaAlaLeuGlyPheMet 180
Db 68585 TTCTCCGCGCAGCGCTTCTCGAAGTTATCCCGCGTTTGAAGCAGCATTAAGTTTATG 68644
Qy 181 LeuProAlaLeuPheMetSerPheLeuLeuAlaSerPheGlnArgLysGlnSerLeuCys 200
Db 68645 CTTCGGGCACCTCTTTATGAGTTTCTGTCTCGCTCTTTCCAGCGCAAACTCTTTGC 68704
Qy 201 ValThrAlaAlaLeuValGlyAlaLeuAlaGlyValThrLeuPheSerIleProValAla 220
Db 68705 GTTACCGCAGCGTTAGTTGGTGGCTTTCAGCGCGTAACGCTATTCTTATTCCTCGCC 68764
Qy 221 IleLeuAlaGlyIleValCysGlyCysLeuThrAlaLeuIleGlnAlaPheTrpGlnGly 240
Db 68765 ATTCTGGCAGCATTTCTGTGGCTGCTCACTGCGTTAATCCAGGCATTTCTGGCAGGA 68824
Qy 241 AlaProAspGluLeu 245
Db 68825 GCGCCCGATGAGCTA 68839

RESULT 12

AE005674_27

WPCOMMENT

Sequence split into 46 fragments LOCUS AE005674 Accession AE005674

Fragment Name	Begin	End
AE005674_00	1	110000
AE005674_01	100001	210000
AE005674_02	200001	310000
AE005674_03	300001	410000
AE005674_04	400001	510000
AE005674_05	500001	610000
AE005674_06	600001	710000
AE005674_07	700001	810000
AE005674_08	800001	910000
AE005674_09	900001	1010000
AE005674_10	1000001	1110000
AE005674_11	1100001	1210000
AE005674_12	1200001	1310000
AE005674_13	1300001	1410000
AE005674_14	1400001	1510000
AE005674_15	1500001	1610000
AE005674_16	1600001	1710000
AE005674_17	1700001	1810000
AE005674_18	1800001	1910000

1900001 2010000
2000001 2110000
2100001 2210000
2200001 2310000
2300001 2410000
2400001 2510000
2500001 2610000
2600001 2710000
2700001 2810000
2800001 2910000
2900001 3010000
3000001 3110000
3100001 3210000
3200001 3310000
3300001 3410000
3400001 3510000
3500001 3610000
3600001 3710000
3700001 3810000
3800001 3910000
3900001 4010000
4000001 4110000
4100001 4210000
4200001 4310000
4300001 4410000
4400001 4510000
4500001 4607203
Continuation (28 of 46) of AE005674 from base 2700001 (AE005674 Shigella flexneri 2a str

Alignment Scores:
Pred. No.: 8.44e-104 Length: 110000
Score: 1243.00 Matches: 243
Percent Similarity: 100.0% Conservative: 2
Best Local Similarity: 99.2% Mismatches: 0
Query Match: 99.6% Indels: 0
DB: 1 Gaps: 0

US-10-073-293A-4 (1-245) x AE005674_27 (1-110000)

Qy 1 MetGluSerProThrProGlnProAlaProGlySerAlaThrPheMetGluGlyCysLys 20
Db 83514 ATGGAAGCCCTACTCACAGCCTGCTCTGGTTCGGCGACCTTCATGGAAGGATGCAAA 83573
Qy 21 AspSerLeuProIleValIleSerTyrIleProValAlaPheAlaPheGlyLeuAsnAla 40
Db 83574 GACAGTTTACCGATTGTTATTAGTTATATTCGGTGGCCTTTGGCTTCGTTCTGNAATCG 83633
Qy 41 ThrArgLeuGlyPheSerProLeuGluSerValPhePheSerCysIleIleTyrAlaGly 60
Db 83634 ACCCGTCTGGGATTCTCTCTCTCGAAAGCGTTTTTTTCTCTCATCATTTATGACGCG 83693
Qy 61 AlaSerGlnPheValIleThrAlaMetLeuAlaAlaGlySerSerLeuTrpIleAlaAla 80
Db 83694 GCGAGCCAGTTTCGTCATTTACCGCATGCTGGCAGCCGGAGTAGTTTGTGGTTGCTGCA 83753
Qy 81 LeuThrValMetAlaMetAspValArgHisValLeuTyrGlyProSerLeuArgSerArg 100
Db 83754 CTGAGCGTCATGCGCAATGAGTTTCGCCCATGTTTGTATGGCCCGTCACTCGGTAGCCGT 83813
Qy 101 IleIleGlnArgLeuGlnLysSerLysThrAlaLeuTrpAlaPheGlyLeuThrAspGlu 120
Db 83814 ATTATTACGCTCTGCAAAATCGAAACCCCTATGGGCGTTTGGCTTCACCGATGAG 83873
Qy 121 ValPheAlaAlaAlaThrAlaLysLeuValArgAsnAsnArgTrpSerGluAsnTrp 140
Db 83874 GTTTTGGCCGCGCACCGCAAACTGCTACGCAATTAATCGCCGCTGGAGCGAGAACTGG 83933
Qy 141 MetIleGlyIleAlaPheSerSerTrpSerTrpValPheGlyThrValIleGlyAla 160
Db 83934 ATGATCGGCATTTGCCCTTCAGTTTCATGGTCACTCGTGGGTATTGTGTACGGTAAATAGGGCA 83993
Qy 161 PheSerGlySerGlyLeuLeuGlnGlyTyrProAlaValGluAlaAlaLeuGlyPheMet 180


```

Db      83994  TTCTCCGCGACGGCTTCTGTCGAGGTTATCCCGCGTTGAAGCTGCATTAGGTTTATG 84053
Qy      181  LeuProAlaLeuPheMetSerPheLeuLeuAlaSerPheGlnArgLysGlnSerLeuCys 200
Db      84054  CTTCCGGCAGCTCTTTATGAGTTTCTCTGCTGCGCTTTTCCAGCGCAACAATCTCTTTCG 84113
Qy      201  ValThrAlaAlaLeuValGlyAlaLeuAlaGlyValThrLeuPheSerIleProValAla 220
Db      84114  GTTACCGCAGCTTTAGTTGGTGGCCCTTGACGCGTAACGCTATTCTTATTCCCGTCGCC 84173
Qy      221  IleLeuAlaGlyIleValCysGlyCysLeuThrAlaLeuIleGlnAlaPheTrpGlnGly 240
Db      84174  ATTTGCGAGCAGCATGCTGTGGCTGCTCACTCGGTATCCAGGCAATCTGCGAAGGA 84233
Qy      241  AlaProAspGluLeu 245
Db      84234  GCGCCCGATGAGCTA 84248

RESULT 13
AE016987 Shigella flexneri 2a str. 2457T DNA linear BCT 22-APR-2003
LOCUS Shigella flexneri 2a str. 2457T section 10 of 16 of the complete
DEFINITION genome.
ACCESSION AE016987 AE014073
VERSION AE016987.1 GI:30042162
SOURCE Shigella flexneri 2a str. 2457T
ORGANISM Shigella flexneri 2a str. 2457T
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Shigella.
REFERENCE 1 (bases 1 to 290380)
AUTHORS Wei,J., Goldberg,M.B., Burland,V., Venkatesan,M.M., Deng,W.,
Fournier,G., Mayhew,G.F., Plunkett,G. III, Rose,D.J., Darling,A.,
Mau,B., Perna,N.T., Payne,S.M., Runyen-Janecky,L.J., Zhou,S.,
Schwartz,D.C. and Blattner,F.R.
Complete Genome Sequence and Comparative Genomics of Shigella
flexneri Serotype 2a Strain 2457T
Infect. Immun. 71 (5), 2775-2786 (2003)
12704152
2 (bases 1 to 290380)
AUTHORS Wei,J., Goldberg,M.B., Burland,V., Venkatesan,M.M., Deng,W.,
Fournier,G., Mayhew,G.F., Plunkett,G. III, Rose,D.J., Darling,A.,
Mau,B., Perna,N.T., Payne,S.M., Runyen-Janecky,L.J., Zhou,S.,
Schwartz,D.C. and Blattner,F.R.
Direct Submission
TITLE Submitted (13-JUN-2002) Genetics Laboratory, University of
JOURNAL Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
PUBMED Location/Qualifiers
REFERENCE 1. 290380
source /organism="Shigella flexneri 2a str. 2457T"
/mol_type="genomic DNA"
/strain="2457T"
/serotype="2a"
/db_xref="taxon:198215"
complement(82..1209)
/gene="gcpE"
/locus_tag="S2733"
complement(82..1209)
/gene="gcpE"
/locus_tag="S2733"
/notes="residues 1 to 372 of 375 are 96.23 pct identical to
residues 1 to 372 of 372 from Escherichia coli K-12 :
B2515"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAPI7888.1"
/db_xref="GI:30042163"
/transl_table=11
/translation="MNNQAPIQRKSTRIYGVNPIGDGAPIAVQSNWTRTRTDVEAT
VNGIKALERGVIVRSVSTMDAAEFKLIKORVNVPLVADIFHDFYRLKVAEYGV
DCLRINPGNIGNERIRNVVDCAKDNIPIRIGVNSAGLEKDIQEKYGPPTQALLES
AMRHVDLDFQKVSVKASDVFLAVESYRLAKIQIDPLHLGITGAGGARSQAV
KSAIGLGLLISEGIGDTRLVSLAADPVEEIKVGFILKLSIRSRGINFIACPTCSRG

```

```

EFVIGTVNALERQLEDIITPMDVSIIGCVVNGPEALVSTLGTGNGKKSGLYEDGV
RKORLDNMDIDOLEARIRAKASQOLDEARRIDVQVGKIIT"
/genes="yfgA"
/locus_tag="S2734"
complement(1236..2249)
/genes="yfgA"
/locus_tag="S2734"
complement(1236..2249)
/genes="yfgA"
/locus_tag="S2734"
/notes="putative membrane; Not classified"
/notes="residues 1 to 337 of 337 are 80.71 pct identical to
residues 1 to 337 of 337 from Escherichia coli K-12 :
B2516"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="AAPI7889.1"
/db_xref="GI:30042164"
/transl_table=11
/translation="MNTTEATHDQNEALTTCGARLNAREQLGSLQQAVERICLKVSTV
RDIEEDKAPADLASTFTLGVIRSYARLVHPEELFLGKQAPLRAAKVAPMQSFSL
GKRKRKRDGMLMTFTLWLVFVGLSGAMWQDHKAQOEITTMADQSSAELSSNSEQ
GQSVPLNTSTTTPATTPASVDVTTATNTOTPAVTAAPADVPQQAQNVVPSQANV
DTAATVPVPTAATTPDGAAPLPTDQAGVTTTAAADPNALVMNFTADCMLEVDATCKLF
SGMRKQGNLNLTGQAPYKUKIGAPAAVQIQYQGVLDLSRFIRTNQVARUTLNAEQS
PAQ"
complement(2534..3688)
/genes="yfgB"
/locus_tag="S2735"
complement(2534..3688)
/genes="yfgB"
/locus_tag="S2735"
/notes="residues 1 to 384 of 384 are 99.47 pct identical to
residues 1 to 384 of 384 from Escherichia coli K-12 :
B2517"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAPI7890.1"
/db_xref="GI:30042165"
/transl_table=11
/translation="MSEQLVTPENVTTKGKINLLDLNQOMBEFFKDLGCKEFTPADQ
VKNWVHYCCDNFDEMTDINKVLRGKKEVAEIRAPEVVEEQSSDGTIKWAIAGDQ
RVETVYIPDDRATLVCSSQVGCALCEKFCSTAGQFNRLRVSIEIIGQVWRAAKIVG
AAKVQQRPTITVVMGSEPLNNVPMAMEIMLDDFGFLSKRRVTLSTGCVIPA
LDKLGMDIVALAISLHAPNDEIRDEIVPINKKYNIEFLAARVRYLEKSNANQGRVT
IEVMLDHNVDGTEHAHOLAELKDKPCKINLIPNPPFPDAPYGRSSNSRIDRFSKVL
MSYGTIVTKTGDDIDACGQLAGDVDRTRKRLKRMQGEAIDIKAV"
complement(3838..4269)
/genes="ndk"
/locus_tag="S2736"
complement(3838..4269)
/genes="ndk"
/locus_tag="S2736"
/function="enzyme; Purine ribonucleotide biosynthesis"
/notes="residues 1 to 143 of 143 are 100.00 pct identical
to residues 1 to 143 of 143 from Escherichia coli K-12 :
B2518"
/codon_start=1
/transl_table=11
/product="nucleoside diphosphate kinase"
/protein_id="AAPI7891.1"
/db_xref="GI:30042166"
/transl_table=11
/translation="MAIERTFSIIKPNVAKNVIGNIFARFEAAGFKIVGTKMLHLTV
EQARGFYABHDGKPFDFGLVEFWMTSGPIVVSVLEGENAVQRHDLGATNPANALACT
LRADYADSLITENGTHGSDSVESNAEIAIFFGEGEVCPRTR"
complement(4418..6517)
/genes="pbpC"
/locus_tag="S2737"
/notes="truncated"
/pseudo
6514..7961
/notes="insertion element"
/insertion_seq="ISSf14"
complement(6534..7730)
/locus_tag="S2738"

```

CDS	complement (6534..7730) /locus_tag="S2738" /function="IS, phage, Tn; Transposon-related functions" /note="residues 1 to 398 of 398 are 90.95 pct identical to residues 1 to 398 from GenPept : >gb AAK18438.1 AF348706.127 (AF348706) putative transposase for IS110 [Shigella flexneri]" /codon_start=1 /transl_table=11 /product="ISSf14 orf" /protein_id="AAP17892.1" /db_xref="GI:30042167" /translation="WTRSSDVESQVFIQVGVGDTHAVAINRSGKRLFDKALPNDB NKLRLSDLKHQGLILLVDQPATIGALPVAVARSEGVLYGLPGLMRRRIADLHAG EAKTVDRAAIIAARLTLPALRRLKLAIEAELSMCGFDDDLAAQTQASNRIR GLLTQHPAPRPELRLEPAVLDDLQRPSPKELASLSEKLLAAQCKLAPRLGKR LAADIAQALAEQTVVFPQTNAAAVLPLRALQLITLRKORDEVALEVEQVLAHPHLYR VLTSMGPVGVTAAARLTTEVACRAFASALAAVAGLAPVTRRSGSSIRGHPRRGN KALKRALFLSAALRDPLSRAYYTRKMSQGRHNQALIALARRRCDVLFAWMRDGTF YTPQGS"
gene	7960..8442 /gene="ssea" /locus_tag="S2740" /note="truncated" /pseudo complement (9261..10037) /gene="sseb" /locus_tag="S2741" complement (9261..10037) /gene="sseb" /locus_tag="S2741" /function="phenotype; Not classified" /note="residues 1 to 258 of 258 are 94.96 pct identical to residues 4 to 261 of 261 from Escherichia coli K-12 : B2522" /codon_start=1 /transl_table=11 /product="enhancer of serine sensitivity" /protein_id="AAP17893.1" /db_xref="GI:30042168" /translation="MSEYTKNELDLEKAAATEPAHRPAFFRTLLLESTVWVPGTAAQGE AVVEDSLDLOHWEKEDQTVIPFTSLAEALQAVEDQAFVMPVTRTFPMTLGETL FLNAKLPTGKFMREISLLIGESNPISLSEGLGELLSVAREPPAQMIDSLTT LFKTIKPVKRAFIKSEKNEEAQPNLLIGIADGDIERIIQATGSVATDTLPGBEPID ICQVKKBGKISHFTHIAPYERRWGGFLRDFKQNRIL" 10137..10170 /note="REP (repetitive extragenic palindromic) element, contains 1 REP sequence" complement (10179..11462) /gene="pepb" /locus_tag="S2742" complement (10179..11462) /gene="pepb" /locus_tag="S2742" /function="putative enzyme; Not classified" /note="residues 1 to 427 of 427 are 92.50 pct identical to residues 30 to 456 of 456 from Escherichia coli K-12 : B2523" /codon_start=1 /transl_table=11 /product="putative peptidase" /protein_id="AAP17894.1" /db_xref="GI:30042169" /translation="MTEAMKITLSTQPADARWGEKATYSINNDGITLHLNAGDILGLI ORAAKIDGLGKHVLQSGEDWADRQWAFQYKPGKTRKVPEDLDLDAQROQLDN RLMIIDWRDITINAPEELGSQLAQRADVLISNVAGDRVTVRTKGDLEQGYMGL HTVGRGERSPVLLALDYNPTGDKAEPYACLVGKITFDSSGYSIKOTAFWMSKSD MGAATVITGALAFATMRGLNKKRKVLFLCCADNLISGNAPFKLGDIIITYKDDVEMNT DABGLVLADGLIDASAQKPEMIIDAATLGAAKTALGNVDHALFDFDALGRLLAS IQDCEATYRKAPVQWAGATGLGVRTINLLTA" complement (11521..11721) /gene="yfhj"
repeat_region	
gene	
CDS	
gene	
CDS	
gene	

/locus_tag="S2743"	
Alignment Scores:	
Pred. No.:	2.84e-103
Score:	1243.00
Percent Similarity:	100.0%
Best Local Similarity:	99.2%
Query Match:	99.6%
Indels:	0
Gaps:	0
DB:	1
US-10-073-293A-4 (1-245) x AE016987 (1-290380)	
Qy	1 MetGluSerProThrProGlnProAlaProGlySerAlaThrPheMetGluGlyCysLys 20
Db	154513 ATGGAAGCCCTACTCCACAGCCTGCTCTGTTGCGGACCTTCATGGAAGGATGCAAA 154572
Qy	21 AspSerLeuProIleValIleSerTyriLeProValAlaPheAlaPheGlyLeuAsnAla 40
Db	154573 GACAGTTTACCGATTGTTATTAGTTATATTCGCGTGCGCTTTGGCTCTCGAATCGG 154632
Qy	41 ThrArgLeuGlyPheSerProLeuGluSerValPhePheSerCysIleIleTyriAlaGly 60
Db	154633 ACCCGTCGGGATTCTCTCCTCTCGAAAGCGTTTTTTCTCCTGCATCATTTATGCAGGC 154692
Qy	61 AlaSerGlnPheValIleThrAlaMetLeuAlaAlaGlySerSerLeuTriPileAlaAla 80
Db	154693 GCGAGCCAGTTCGTTCATTACCGCGATGCTGCACGCCGGAGTAGTTTGTGGTGTCTGCA 154752
Qy	81 LeuThrValMetAlaMetAspValArgHisValLeuTyriGlyProSerLeuArgSerArg 100
Db	154753 CTGAGCGTCATGGCAATGGATGTTTCGTCATGTTTGTATGCGCGCTCACTCGTAGCCGT 154812
Qy	101 IleIleGlnArgLeuGlnLysSerIysThrAlaLeuTriPileAlaPheGlyLeuThrAspGlu 120
Db	154813 ATTATTACGCGCTCTGCAAAATCGAAAACCCCTATGCGGCGTTTGGCCTCAGCGATGAG 154872
Qy	121 ValPheAlaAlaAlaThrAlaLysLeuValArgAsnAsnArgArgTrpSerGluAsnTrp 140
Db	154873 GTTTTTCGCGCGCGCACCGCAAAACTGTAGCGCAATAATCGCCGCTGAGCGAGAACTGG 154932
Qy	141 MetIleGlyIleAlaPheSerSerTrpSerSerTrpValPheGlyThrValIleGlyAla 160
Db	154933 ATGATCGGCAATTCGCTTCAGTTTCATGTCATCGTGGGTATTTGGTACGGTAATAGGGCA 154992
Qy	161 PheSerGlySerGlyLeuLeuGlnGlyTyriProAlaValGluAlaAlaLeuGlyPheMet 180
Db	154993 TTCTCCGCGACGCGCTTCTGCAAGGTTATCCCGCGCTTGAAGCTGCATTAGGTTTATG 155052
Qy	181 LeuProAlaLeuPheMetSerPheLeuLeuAlaSerPheGlnArgLysGlnSerLeuCys 200
Db	155053 CTTCCGGCACCTCTTTATGAGTTTCTGCTCGCCTCTTTCAGCGCAACAACTCTTTTGC 155112
Qy	201 ValThrAlaAlaLeuValGlyAlaLeuAlaGlyValThrLeuPheSerIleProValAla 220
Db	155113 GTTACCGCAGCTTTAGTTGGTGCCCTTCGACGGCGTAACGCTATTCTTATTCCTCCGTCGCC 155172
Qy	221 IleLeuAlaGlyIleValCysGlyCysLeuThrAlaLeuIleGlnAlaPheTrpGlnGly 240
Db	155173 ATTCGGCAGCATTTGCTGTGGCTGCTCACTGCGTTAATTCACGGCATTCTGGCAAGGA 155232
Qy	241 AlaProAspGluLeu 245
Db	155233 GCGCCCGATGAGCTA 155247
RESULT 14	
AR386287	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
Sequence 3016 from patent US 6610836.	
756 bp DNA linear PAT 18-DEB-2003	
Unknown.	
Unknown.	

Unclassified.
1 (bases 1 to 756)
Breton,G.L. and Osborne,M.
Nucleic acid amino acid sequences relating to Klebsiella pneumoniae
for diagnostics and therapeutics
Patent: US 6510836-A 3016 26-AUG-2003;
Genome Therapeutics Corporation; Waltham, MA
Location/Qualifiers
1 .756
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 4.06e-85 Length: 756
Score: 1013.50 Matches: 200
Percent Similarity: 88.3% Conservative: 18
Best Local Similarity: 81.0% Mismatches: 26
Query Match: 81.2% Indels: 3
DB: 6 Gaps: 1
US-10-073-293A-4 (1-245) x AR386287 (1-756)
Qy 1 MetGluSerProThrProGlnProAla-----ProGlySerAlaThrPheMetGlu 17
|||:|||||
Db 10 ATGGAAACCCCGCGCCCTTAACCTGTGCCCTTACCTGAACCGCTGCCACCGTCGGCGAA 69
|||:|||||
Qy 18 GlyCysIysAspSerLeuProIleValIleSerTyrlleProValAlaPheAlaPheGly 37
|||:|||||
Db 70 GCGTCGCAAGACAGTCTCCCATCGTCATCAGTACCTTCGCGTGGCATTCGCCCTTCGGA 129
|||:|||||
Qy 38 LeuAenAlaThrArgLeuGlyPheSerProLeuGluSerValPhePheSerCysIleIle 57
|||:|||||
Db 130 CTCAATGCTACCGCGTGGCTTCACCCCGCTGGAAGCCCTGTTCTTCTCTCGCATATT 189
|||:|||||
Qy 58 TyrAlaGlyAlaSerGlnPheValIleThrAlaMetLeuAlaAlaGlySerSerLeuTyr 77
|||:|||||
Db 190 TACCGCGCGCGCAGCGATTTGTGATCACCGCCATGTTGGCGCGCGCGATTCGCTGTGG 249
|||:|||||
Qy 78 IleAlaAlaLeuThrValMetAlaMetAspValArgHisValLeuTyrGlyProSerLeu 97
|||:|||||
Db 250 GTCGCGCGCCCTCAGCGTGATGGATGATGATGATGATGATGATGATGATGATGATGATG 309
|||:|||||
Qy 98 ArgSerArgIleIleGlnArgLeuGlnIysSerIysThrAlaLeuTyrAlaPheGlyLeu 117
|||:|||||
Db 310 CGCAGCGCATTCGCTCAGCGCTCGATATAAAGAAACCGCCCTCTCGGCTTCGGCTG 369
|||:|||||
Qy 118 ThrAspGluValPheAlaAlaThrAlaLysLeuValArgAsnAsnArgArgTyrSer 137
|||:|||||
Db 370 ACCGATGAGTTTTCGCGCGCGCTACCGCGCGCTGTTTCGCGATACCGCGCTGGAGC 429
|||:|||||
Qy 138 GluAenTyrMetIleGlyIleAlaPheSerSerTyrPheSerTyrValPheGlyThrVal 157
|||:|||||
Db 430 GAAACGTGATGCTCGGCTCGCTTTTACCTCGTGGGCATCGTGGGTGTCGGCAGCCGCT 489
|||:|||||
Qy 158 IleGlyAlaPheSerGlySerGlyLeuLeuGlnGlyTyrProAlaValGluAlaAlaLeu 177
|||:|||||
Db 490 GCCGCGCGCTGTCGCGCAACGGCTGCTGTGGATTATCCGCGGTGGAGCGCGCCCTC 549
|||:|||||
Qy 178 GlyPheMetLeuProAlaLeuPheMetSerPheLeuLeuAlaSerPheGlnArgGlyGln 197
|||:|||||
Db 550 GGCTTTATGTCGCGCGCGCTGTTTATGATGCTTCTGCTGGCTCTTTCACGCGTCAGCAG 609
|||:|||||
Qy 198 SerLeuCysValThrAlaAlaLeuValGlyAlaLeuAlaGlyValThrLeuPheSerIle 217
|||:|||||
Db 610 TCATCTGCGTCAGCGCGCGCTGCGCGCGCGCTGCGCGCGCATTCGTATTATTTTCATT 669
|||:|||||
Qy 218 ProValAlaIleLeuAlaGlyIleValCysGlyCysLeuThrAlaLeuIleGlnAlaPhe 237
|||:|||||
Db 670 CCGCGCGGATCTTCGCGGAAATGTCGCGGATGTCGCGGATGTCGCGGCTGCTTCAGGCCATG 729
|||:|||||
Qy 238 TrpGlnGlyAlaProAspGlu 244
|||:|||||

Db 730 CTCAAGGGGATGCCGATGAA 750
RESULT 15
AE013695/c
LOCUS
Yersinia pestis KIM section 95 of 415 of the complete genome.
ACCESSION AE013695 AE009952
VERSION AE013695.1 GI:21957652
KEYWORDS
SOURCE
Yersinia pestis KIM
Yersinia pestis KIM
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Yersinia.
1 (bases 1 to 10733)
Deng,W., Burland,V., Plunkett,G. III, Boutin,A., Mayhew,G.F.,
Liss,P., Perna,N.T., Rose,D.J., Mau,B., Zhou,S., Schwartz,D.C.,
Fetherston,J.D., Lindler,L.E., Brubaker,R.R., Plana,G.V.,
Straley,S.C., McDonough,K.A., Nilles,M.L., Matson,J.S.,
Blattner,F.R. and Perry,R.D.
Genome Sequence of Yersinia pestis KIM
J. Bacteriol. 184 (16), 4601-4611 (2002)
12142430
2 (bases 1 to 10733)
Deng,W., Burland,V., Plunkett,G. III, Boutin,A., Mayhew,G.F.,
Liss,P., Perna,N.T., Rose,D.J., Mau,B., Zhou,S., Schwartz,D.C.,
Fetherston,J.D., Lindler,L.E., Brubaker,R.R., Plana,G.V.,
Straley,S.C., McDonough,K.A., Nilles,M.L., Matson,J.S.,
Blattner,F.R. and Perry,R.D.
Direct Submission
Submitted (21-FEB-2002) Genetics, University of Wisconsin, 445
Henry Mall, Madison, WI 53706, USA
Location/Qualifiers
1 .10733
/organism="Yersinia pestis KIM"
/mol_type="genomic DNA"
/strain="KIM"
/db_xref="taxon:187410"
complement (117. .2759)
/gene="y0917"
complement (117. .2759)
/gene="y0917"
/notes="Residues 1 to 880 of 880 are 75.45 pct identical to
residues 1 to 880 of 880 from E. coli K12 : B2584;
residues 1 to 880 are 75.68 pct identical to
residues 1 to 880 of 886 from GenPept :
>gb|AA57701.1|AE005488.12 (AE005488) orf, hypothetical
protein [Escherichia coli O157:H7 EDL933]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA084499.1"
/db_xref="GI:21957653"
/translation="MSQRGLEALLRPKSTAVIGASKEPAGERFLMRLNDGGFNGPI
LPVTPNHKAVCGVLAIVANIASLEFIPDLALCTHDCNLTLLLEDLGNRGCAAILISA
APQDFPELCAQRMHRLGNSLGLAPWQGNASFSVPVKKGLAFISQSAVA
NTLLDWAQREVGFSYFIALGSDSLDIDVDLDFLARDSTSAIMLYIEHISDARFL
SASRSARNKPLVVKSGRSQAQQLNGQDAAAYDAAIQAGILRVODTHLFLFA
VETLSHMRPLRGERLLIVNSGSAAPAMALDELIRNRKGLATLSDATQSALSEALPPFV
ALRNPDLRDDASAEYLAALKPLDSSDYDTLLIHSPSAAPGCKTALLISAIQ
HPGKRITLLTNWCGYSQDARLFTEAGIPTRTPEGALTAFMNMVETRRNQKQK
ETPALPIGTANTAHQILQRLAEGATQLDTHEVQILEAYGLRTLPWTIASDSVE
AHVIAERGLGPVAIKLRSPDIPIHKSVEQVGMVLYRTAIEVQRAADDILDRVKTPOA
RHGLLVQSMANRAGAQLRIAVEQDAIFGLIMLGGGIEWHETQAAVALPPLNV
LARVLIQAVKGGKIRSGSLQPLDIPGLSRLLAVOVSNLIDCPETRLRIDHPVLASG
SEFTLDVSNQLAPVTDQPOARLAIRPYHELEKVTLRDNSQCLFRPILPEPEPLK
LFIDQVTKEDLYTFSEINEFSDHDLNMTQIDYDREMAFVAVRQNSEPGEIIGVTR
AFSDPNIDAEFAVLRSLDKGLGLRALLERKIRYARSHGLSRLTAVTPNNRMTIG
LAQKGFITDVIQEDGIVNLELT"
complement (2802. .3551)
/gene="y0918"
complement (2802. .3551)
/gene="y0918"
/notes="residues 14 to 248 of 249 are 60.85 pct identical

```

to residues 7 to 239 of 240 from E. coli K12 : B2583"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAM84500.1"
/db_xref="GI:21957654"
/translation="MTDPTFTTGHPTLPTAITTENAVLRLOORLAAATRAYRARGCRAI
RCQCLLAERECLODTIKPOQANSRFLIMEDTPKLPSTNGRLIADILPPTHAPLWA
RTEVDPPELLAAISPLRQPYVFFEAYAEPRQVINQLPISDKPEPLFILDGTWTEAK
RMRKSPYLQAGLLSLQVSHLSYQUREAQRPBOHCTVEVATALLHQAGDIOAEGGL
RDFHYFQQYLAGKPHRPVGRVTAASHEENA"
complement(3721..4158)
/gene="trxC"
/feature="synonym: y0919"
complement(3721..4158)
/gene="trxC"
/function="putative"
/feature="residues 1 to 139 of 145 are 77.69 pct identical to
residues 1 to 139 of 139 from E. coli K12 : B2582"
/codon_start=1
/transl_table=11
/product="putative thioredoxin-like protein"
/protein_id="AAM84501.1"
/db_xref="GI:21957655"
/translation="MNTVCTACMATNRLPEERIDDGKCGRCHSLPFGVEVINATAET
LDKLLQDLPWIDFWPWCPCPSFAPIPAEATAERAGKVRPVKYNTEAPALSTRP
RIRSIPTIMLRNGKIMDLNGLGAVFKAPFDNLDQLSRDPSN"
4395..5576
/gene="y0920"
4395..5576
/gene="y0920"
/feature="residues 6 to 391 of 393 are 60.05 pct identical to
residues 1 to 345 of 345 from E. coli K12 : B2581;
residues 1 to 389 of 393 are 65.72 pct identical to
residues 2 to 364 of 366 from GenPept : >emb|CAC44352.1|
(AJ410308) YfiP protein [Pectobacterium chrysanthemi]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAM84502.1"
/db_xref="GI:21957656"
/translation="MVEKTMDSFGKNGKVKVMYRSDSSDDRNSNRDRNNSRR
DNSNRDNNRGRSOPRPNAGRPGDNROGDNTRQNA PRNNDSSRRPDGSRNERDRPSR
PARSEGSPYSPKTVTSRPSRSEPEPDHGIGISKSHIDPAQLRQRAETRVYGENA
CKALFESRPAIVRAFWFQSVTPRFRALKMAANRKAHYHVESDELAASGTEHHGG
VCFILIKRGQDAETYLQQQQAQKDCVLAKEVGNPHNLGIMTCAHFNGVGLLOD
PAMLESAAVTAEGGAHHIKAINADDFLSVLDTRFKAGYTIIVTSSHKGVSLKAEI
PAKMVVLQGESDGLTDSAWQQGDLSSVIGTGVRVESLNVSVATGILLAEWRRQNSAQ
G"
complement(5699..7234)
/gene="emrB"
/feature="synonym: y0921"
complement(5699..7234)
/gene="emrB"
/function="putative transport; drug/analog sensitivity"
/feature="multidrug resistance; residues 3 to 511 of 511 are
84.08 pct identical to residues 4 to 512 of 512 from E.
coli K12 : B2686"
/codon_start=1
/transl_table=11
/product="putative membrane translocase"
/protein_id="AAM84503.1"
/db_xref="GI:21957657"
/translation="MAQPLEGAPLAWMTVALSLATFMQVILDSTIANVAIPTIADGLG
SSNSQGTWIVISFGVNAISIPVTGMLAKRGEVRLFLWSTGLFVLASWLCGMSNLG
MLIFPRVQIGVAGVLIPLSQSLILNNYPPAKRSMALAWMTIVVAPIFGPIIGYVI
SDNYHWGFIPIINIGLVVLLAGSTLKGRETKEIRPIDTIGLVLLVIGIGLQIM
LDQKELDFNSTEILVTVAVVAITFLIWEIITDHPVIDLSLFRNRTIICLCL
SLAYMLYFGAIVLPQLLQVEYGYTATWAGLASAPDGLPVLLSPLIFGRFAHRLDMQI
LVTFISFYAVCFYWRVTFEFGDASAPQFFQGFPAIACFFMPLTITLSGLPPE
RMAAASLSNFWRTLASIGTSITITLWTORESLHQAOLTEFVFNPNPNTAQYQELE
KLGMSQQQASAVLAREITNOGLIISANEIPLWSAGVPLMLLALIWFARPPPTSGGGG
GAH"

```

```

gene
complement(7273..8457)
/gene="emrA"
/feature="synonym: y0922"
complement(7273..8457)
/gene="emrA"
/function="transport; drug/analog sensitivity"
/feature="residues 5 to 393 of 394 are 69.66 pct identical to
residues 1 to 389 of 390 from E. coli K12 : B2685;
residues 5 to 393 of 394 are 71.20 pct identical to
residues 1 to 389 of 390 from GenPept : >gb|AAU21699.1|
(A808828) multidrug resistance secretion protein
[Salmonella typhimurium LT2]"
/codon_start=1
/transl_table=11
/product="multidrug resistance secretion protein"
/protein_id="AAM84504.1"
/db_xref="GI:21957658"
/translation="MENTSESAGAQAPQRPQNKQKRVLILLTIFIMIGVAYILI
YFVLVRHQBTDNAVISGNQVIMSOVPGSVSVSHFENDTFVKSGDVLVLDPTDAE
QAEQAKTALANSVRQTHQLILNSKOYQANIALKKTLSQANDLKRVLGRAAVIG
REELQHARDAVEAQAQSLMAVQOYNANQALVNLPLEKQPAEQAAKWRDAWLTQ
RTKVSPISGTVSVSRSSVQVGAIESSGTPLMVAVPADQIWDANFEKQTOLANRMLGQQA
TIVTDFYGDVVYQGVGLMDGTSGSFLPQAQNTGNMIVKVRQLPRALDEKQL
KEHPLRTIGLSLVKVDNTANTDGHVLAQNVRKEPAFMTNALSLEDPAPVNIISDIHAN
AD"
complement(8829..9365)
/gene="emrR"
/feature="synonym: y0923"
complement(8829..9365)
/gene="emrR"
/function="regulator; plasmid-related functions"
/feature="residues 9 to 174 of 178 are 80.12 pct identical to
residues 1 to 166 of 176 from E. coli K12 : B2684"
/codon_start=1
/transl_table=11
/product="regulator of plasmid mcrB operon (microcin B17
synthesis)"
/protein_id="AAM84505.1"
/db_xref="GI:21957659"
/translation="MKLMRRLLMESSPSPTEQMINSRAKQKQFPYQEIILLTRLSMHEM
LSKILLENRMKLQAGINETLFWALITLDAGRSIQPSLSAALGSRSTNTRIADIE
LEKKGWTERRSHNDRRCLHLTEAGIEFLQLLPPQHKCLHPLWSTLTREQOQOLE
MLTRKLLTRLDQWEISBQ"
complement(9513..9854)
/gene="y0924"
complement(9513..9854)
/gene="y0924"
/feature="residues 1 to 112 of 113 are 54.38 pct identical to
residues 1 to 111 of 111 from E. coli K12 : B2683;
residues 1 to 107 of 113 are 66.35 pct identical to
residues 1 to 107 of 113 from GenPept : >emb|CAC44348.1|
(AJ410307) YgaH protein [Pectobacterium chrysanthemi]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAM84506.1"
/db_xref="GI:21957660"
/translation="MNTDILMIGLVGTVNYLFRYLPFRFGPARKQGTGLLRGRMSLLL

```

Alignment Scores:

Pred. No.:	1.91e-70	Length:	10733
Score:	871.50	Matches:	168
Percent Similarity:	83.1%	Conservative:	29
Best Local Similarity:	70.9%	Mismatches:	35
Query Match:	69.8%	Indels:	5
DB:	1	Gaps:	1

US-10-073-293A-4 (1-245) x AE013695 (1-10733)

Qy 4 ProThrProGlnProAlaProGlySer-----AlaThrPheMetGluGly 18

Db 10621 CCAAAACCAACCAACGACCCCTTCTTCTACAACTTCAGCGCGCGAGCTTCGTTGAAGGT 10562

```
Qy      19  CysLysAspSerLeuProIleValIleSerTyrIleProValAlaPheAlaPheGlyLeu 38
Db      10561  ATTACGGATAGCCTGCCCATTTGTTATTCGGCTATTACCGTCGCTTTGCGCTTTGGCCTC 10502

Qy      39  AsnAlaThrArgLeuGlyPheSerProLeuGluSerValPhePheSerCysIleIleTyr 58
Db      10501  AGTCGGTAAACCTGGGTTTTACCCCGCTGGAAGCATTTTCTTTCTGCAATTATTAT 10442

Qy      59  AlaGlyAlaSerGlnPheValIleThrAlaMetLeuAlaAlaGlySerSerLeuTyrIle 78
Db      10441  GCTGGTGCCAGCCCAATTGTGATTACCGCCCTGCTGAGCGCGGGATGTCATTGTGGGTA 10382

Qy      79  AlaAlaLeuThrValMetAlaMetAspValArgHisValLeuTyrGlyProSerLeuArg 98
Db      10381  TCGGCACCTGACGGTGATGGCAATGGATGCGCCATATTTGTATGCGCCGACATTAAA 10322

Qy      99  SerArgIleIleGlnArgLeuGlnLysSerLysThrAlaLeuTyrAlaPheGlyLeuThr 118
Db      10321  CACCGCATTTTGGCAAACTATACAGGTAAATAAAACGGCCTTTATGGGCTTTTGGTTAAAC 10262

Qy      119  AspGluValPheAlaAlaAlaThrAlaLysLeuValArgAsnAsnArgArgTyrSerGlu 138
Db      10261  GATGAAGTGTTCGGCTGCAACCACTAAATTTGATGAAGGATCAACGGCGCTGGAGCGAG 10202

Qy      139  AsnTrpMetIleGlyIleAlaPheSerSerTrpSerSerTrpValPheGlyThrValIle 158
Db      10201  AACTGGATGCTCGGTATCGCAGTCACCTCGTGGTTATCTTGGGTCTTGGGAACGGCCATC 10142

Qy      159  GlyAlaPheSerGlySerGlyLeuGlnGlyTyrProAlaValGluAlaAlaLeuGly 178
Db      10141  GCGCGGATGTCGGTAATGGGCCCTTGAGAGAACTACCCCGCATTTGAAGCGTCCCTCTCC 10082

Qy      179  PheMetLeuProAlaLeuPheMetSerPheLeuLeuAlaSerPheGlnArgLysGlnSer 198
Db      10081  TTTATGCTGCTGCATTATTCCTCAGTTTTTTGTAGCTGTTTAAACGCCAATACAGC 10022

Qy      199  LeuCysValThrAlaAlaLeuValGlyAlaLeuAlaGlyValThrLeuPheSerIlePro 218
Db      10021  CTGACGGTGATCGCTTCGCTTTCAGGTGCCTTATTAGGGGTTTTATTGTTCTCAATTCCA 9962

Qy      219  ValAlaIleLeuAlaGlyIleValCysGlyCysLeuThrAlaLeuIleGln 235
Db      9961  GTGGCTATTTGGCCCGGATTCGGGTGGTGTCTCGCGACTTTGCTTCAG 9911
```

Search completed: February 16, 2006, 19:45:57
Job time : 6383.4 secs

This Page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 16, 2006, 17:00:29 ; Search time 737.065 Seconds
(without alignments)
2215.341 Million cell updates/sec

Title: US-10-073-293A-4

Perfect score: 1248

Sequence: 1 MESPTQPAPGSAFTMEGCK.....VCGCLTALIOAFWQGAPDEL 245

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/abs/ABSSWEB spool/US10073293/runat_15022006_120733_951/app_query.fasta.1
-DB=Geneseq -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blom62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs07
-USER=US10073293 @CGN 1.1 900 @runat_15022006_120733_951 -NCPU=6 -ICPU=3
-NO MMAP -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N Geneseq 21.*
1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002as.*
7: Geneseqn2002bs.*
8: Geneseqn2003as.*
9: Geneseqn2003bs.*
10: Geneseqn2003cs.*
11: Geneseqn2003ds.*
12: Geneseqn2004as.*
13: Geneseqn2004bs.*
14: Geneseqn2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1248	100.0	738	6	ABV75656 E. coli L
2	1248	100.0	738	14	ADZ39950
3	1248	100.0	738	14	ADZ36194 Escherich
4	1013.5	81.2	756	11	ACH97221 Klebsiell

5	841.5	67.4	759	10	ACF69125
6	841.5	67.4	110000	10	ACF67367 20
7	841.5	67.4	243072	10	ACF65382
8	592	47.4	801	9	ADA31112 DNA encod
9	426	34.1	1589	4	AAI35397
10	426	34.1	1589	4	AAI62810 Human mus
11	426	34.1	1589	4	ADK43465 DNA encod
12	426	34.1	1589	5	ADM19434 Novel hum
13	426	34.1	1589	8	ABX58385 cDNA enco
14	426	34.1	1589	12	ADI53852
15	426	34.1	1589	12	ADJ28112 Human mus
16	426	34.1	1595	4	AAI35674 Human mus
17	426	34.1	1595	4	ADK43782 DNA encod
18	426	34.1	1595	5	ADM19672 Novel hum
19	426	34.1	1595	8	ABX58662 cDNA enco
20	426	34.1	1595	12	ADI54169 cDNA enco
21	426	34.1	1595	12	ADJ28389 Human mus
22	290.5	23.3	4947	2	AAQ86927 Tspel17 po
23	246.5	19.8	744	11	ACH95512 Klebsiell
24	238.5	19.1	12445	2	AAI31179 Enterococ
25	238.5	19.1	12445	6	ABX98974 Enterococ
26	229.5	18.4	693	4	AAH52437 S. epider
27	229.5	18.4	3058	4	AAH54221 S. epider
28	229.5	18.4	3077	4	AAH54230 S. epider
29	228	18.3	110000	6	ABA03041 14
30	225.5	18.1	729	6	ABN92854
31	225.5	18.1	729	13	ADSO3654
32	225	18.0	110000	6	ABQ69245 14
33	225	18.0	110000	6	ABQ67195 2
34	223	17.9	6157	6	ABQ70939
35	220.5	17.7	19760	13	ADV87714 Streptoco
36	220.5	17.7	19760	13	ADV78967 Streptoco
37	220.5	17.7	110000	6	ABN71527 19
38	220.5	17.7	110000	13	ADV81204 20
39	219.5	17.6	690	6	ABN71161 Streptoco
40	219.5	17.6	693	6	ABN69193 Streptoco
41	219.5	17.6	693	13	ADV83888 Streptoco
42	214.5	17.2	768	10	ADC91639 E. faeciu
43	210.5	16.9	1038	11	ABD12772 Pseudomon
44	210.5	16.9	1086	11	ABD12814 Pseudomon
45	204.5	16.4	714	8	ACF74375 Staphyloc

ALIGNMENTS

RESULT 1
ABV75656
ID ABV75656 standard; DNA; 738 BP.
XX
XX ABV75656;
AC
XX
DT 29-JAN-2003 (first entry)
XX
DE E. coli L-amino acid producing gene #1.
XX
XX L-amino acid; gene; ss; E.coli.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT CDS 1..738
FT /*tag= a
XX
PN EPI239041-A2.
XX
PD 11-SEP-2002.
XX
PF 13-FEB-2002; 2002EP-00003335.
XX
PR 13-FEB-2001; 2001RU-00103865.
PR 26-FEB-2001; 2001RU-00104998.
PR 26-FEB-2001; 2001RU-00104999.
PR 28-JUN-2001; 2001RU-00117632.

PR 28-JUN-2001; 2001RU-00117633.
 XX (AJIN) AJINOMOTO CO INC.
 XX Tabolina EA, Rybak KV, Khourges EM, Voroshilova EB, Gussyatiner MM;
 XX P-PSDB; ABP57756.
 DR WPI; 2002-699856/76.
 XX Novel L-amino acid producing Escherichia bacterium, is modified to
 PT enhance L-amino acid production by enhancing the activities of protein
 PT capable of making bacterium to have enhanced resistance to L-amino acids.
 XX Claim 1; Page 21-22; 33pp; English.
 XX The invention relates to a novel L-amino acid producing bacterium of
 CC genus Escherichia, modified to enhance L-amino acid production by
 CC enhancing the activities of a protein. The novel bacterium is useful for
 CC producing L-amino acid e.g. L-Thr, L-Val, L-Pro, L-Leu, L-Met and L-Arg,
 CC by cultivating the bacterium in a culture medium and collecting L-amino
 CC acid to be produced and accumulated from the culture medium. The present
 CC sequence represents a gene of the invention which causes increased L-
 CC amino acid production in E. coli
 XX SQ Sequence 738 BP; 136 A; 192 C; 198 G; 212 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 8.45e-124 Length: 738
 Score: 1248.00 Matches: 245
 Percent Similarity: 100.0% Conservativity: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 100.0% Indels: 0
 DB: 6 Gaps:
 US-10-073-293A-4 (1-245) x ABV75656 (1-738)
 QY 1 MetGluSerProThrProGlnProAlaProGlySerAlaThrPheMetGluGlyCysLys 20
 Db 1 ATGGAAGCCCTACTCCACAGCTCTCTGTTCCGGACCTTCATGAAGGATGCAAA 60
 QY 21 AspSerLeuProIleValIleSerTyrIleProValAlaPheAlaPheGlyLeuAsnAla 40
 Db 61 GACAGTTTACCGATTGTTATTAGTTATATCCGGTGGCTTTCGGTCTCGAATCG 120
 QY 41 ThrArgLeuGlyPheSerProLeuGluSerValPhePheSerCysIleIleTyrAlaGly 60
 Db 121 ACCCGTCTGGGATTCTCTCTCGAAAGCGTTTTTTTCTCTGTCATCAITTTATGCGAGC 180
 QY 61 AlaSerGlnPheValIleThrAlaMetLeuAlaAlaGlySerSerLeuTrpIleAlaAla 80
 Db 181 GCGAGCCAGTTCTGTCATTTACCGCATGCTGGCAGCCGGAGTAGTTTGTGGATTGCTGCA 240
 QY 81 LeuThrValMetAlaMetAspValArgHisValLeuTyrGlyProSerLeuArgSerArg 100
 Db 241 CTGACCGTCATGCAATGGATGTTCCGCATGTGTTGTATGGCCGTCACGTAGCCGT 300
 QY 101 IleIleGlnArgLeuGlnIleSerIleThrAlaLeuTrpAlaPheGlyLeuThrAspGlu 120
 Db 301 ATTATTACGCGTCTGCAAAATCGAAACCCGCTGTGGGGGTTTGGCGCTGACCGATGAG 360
 QY 121 ValPheAlaAlaThrAlaIleValLeuValArgAsnAlaArgArgArgArgArgArgArg 140
 Db 361 GTTTTGGCCGCGCAACCGCAAACTGGTAGCGCAATATCGCCGCTGGAGCGAGAACATGG 420
 QY 141 MetIleGlyIleAlaPheSerSerTrpSerTrpValPheGlyThrValIleGlyAla 160
 Db 421 ATGATCGCATGTCCTTCAGTTTCATGTCATGTCGTTATTTGGTACGTAATAGGGCA 480
 QY 161 PheSerGlySerGlyLeuLeuGlnGlyTyrProAlaValGluAlaAlaLeuGlyPheMet 180
 Db 481 TTCTCCGCGAGCGGCTGTGTCAGAGTTATCCCGCGTTGAAGCTGTCATTAGTTTATG 540
 QY 181 LeuProAlaLeuPheMetSerPheLeuLeuAlaSerPheGlnArgLysGlnSerLeuCys 200

Db 541 CTTCGGGCACTCTTTATGAGTTTCCTGCTCGCTCTTTCCAGCGCAACATCTCTTGC 600
 QY 201 ValThrAlaAlaLeuValGlyAlaLeuAlaGlyValThrLeuPheSerIleProValAla 220
 Db 601 GTTACCGCAGCGTTAGTTGGTGGCTTTCGAGCGCTAAACGCTATTCTTATTCGCGTCC 660
 QY 221 IleLeuAlaGlyIleValCysGlyCysLeuThrAlaLeuIleGlnAlaPheTrpGlnGly 240
 Db 661 ATTCTGGCAGGCAATGTCTGTGGCTGCTCTACTCGCTTAATCCAGGCATCTCGCAAGGA 720
 QY 241 AlaProAspGluLeu 245
 Db 721 GCGCCCGATGAGCTA 735
 RESULT 2
 ID ADZ39950 standard; cDNA; 738 BP.
 XX AC ADZ39950;
 XX 14-JUL-2005 (first entry)
 DE E.coli ygaZ gene for improved amino acid production in bacteria.
 XX amino acid production; fermentation; transport protein; gene; ss.
 XX Escherichia coli.
 OS Location/Qualifiers
 FH 1..738
 FT /*tag= a
 FT /product= "amino acid biosynthesis protein"
 FT /function= "putative transport protein"
 FT /gene= "ygaZ"
 XX EP1526179-A1.
 PD 27-APR-2005.
 XX 13-FEB-2002; 2004EP-00028876.
 XX 13-FEB-2001; 2001RU-00103865.
 PR 26-FEB-2001; 2001RU-00104998.
 PR 26-FEB-2001; 2001RU-00104999.
 PR 28-JUN-2001; 2001RU-00117632.
 PR 28-JUN-2001; 2001RU-00117633.
 PR 13-FEB-2002; 2002EP-00003335.
 XX (AJIN) AJINOMOTO CO INC.
 XX Tabolina EA, Rybak KV, Khourges EM, Voroshilova EB, Gussyatiner MM;
 XX WPI; 2005-308098/32.
 DR P-PSDB; ADZ39951.
 DR GENBANK; U00096.
 XX New L-amino acid producing bacterium belonging to the genus Escherichia,
 PT useful for producing L-amino acids, e.g. L-threonine, L-valine, L-
 PT proline, L-methionine, or L-arginine.
 XX Disclosure; SEQ ID NO 3; 35pp; English.
 XX The invention relates to an L-amino acid producing bacterium belonging to
 CC the genus Escherichia, where the bacterium has been modified so that the
 CC L-amino acid production by the bacterium is enhanced by enhancing
 CC activities of proteins by transformation of the bacterium with DNA coding
 CC for protein or by alteration of an expression regulation sequence of the
 CC DNA on the chromosome of the bacterium. The L-amino acid producing
 CC bacterium is useful for producing L-amino acids by fermentation,
 CC including L-threonine, L-valine, L-proline, L-methionine, or L-arginine.
 CC The genes are useful for improving L-amino acid productivity. This
 CC sequence corresponds to the ygaZ gene encoding one of the L-amino acid

CC biosynthesis pathway proteins. The protein is a putative transport
 CC protein within the biosynthesis pathway.

SQ Sequence 738 BP; 136 A; 192 C; 198 G; 212 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 8.45e-124 Length: 738
 Score: 1248.00 Matches: 245
 Percent Similarity: 100.0% Conservatives: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 100.0% Indels: 0
 DB: 14 Gaps: 0

US-10-073-293A-4 (1-245) x AD239950 (1-738)

Qy 1 MetGluSerProThrProGlnProAlaProGlySerAlaThrPheMetGluGlyCysLys 20
 Db 1 ATGGAAGCCCTACTCCACACCTGCTCTGGTTCGGCGACCTTCATGGAAGGATGCAAA 60
 Qy 21 AspSerLeuProIleValIleSerTyrIleProValAlaPheAlaPheGlyLeuAsnAla 40
 Db 61 GACAGTTTACCGATTGTTATTAGTTATATTCGGGTGGCTTTGGTTCGGTCTGAATGCG 120
 Qy 41 ThrArgLeuGlyPheSerProLeuGluSerValPhePheSerCysIleIleTyrAlaGly 60
 Db 121 ACCCGTCTGGGATTCTCTCCTCTCGAAAGCGTTTTTTTCTCTGCATCATTTATGCGAGC 180
 Qy 61 AlaSerGlnPheValIleThrAlaMetLeuAlaAlaGlySerSerLeuTrpIleAlaAla 80
 Db 181 GCGAGCCAGTTTCGTCATTACCGCGATCTGCGACCGGAGTAGTTTGTGGATTGCTGCA 240
 Qy 81 LeuThrValMetAlaMetAspValArgHisValLeuTyrGlyProSerLeuArgSerArg 100
 Db 241 CTGACCGTCATGGCAATGGATGTTCCGCACTGTGTGTATGTCGCCGTCACTGCGTAGCGGT 300
 Qy 101 IleIleGlnArgLeuGlnLysSerIysThrAlaLeuTrpAlaPheGlyLeuThrAspGlu 120
 Db 301 ATTATTACGCGCTCTGCAAAATCGAAACCGCCCTGTGGCGTGTGGCTGACGATGAG 360
 Qy 121 ValPheAlaAlaThrAlaLysLeuValArgAsnAenArgTrpSerGluAsnTrp 140
 Db 361 GTTTTTCGCGCGCCGACCGCAAACTGTGTACGCAATATCCCGCTCGAGCGAGACTGG 420
 Qy 141 MetIleGlyIleAlaPheSerSerTrpSerTrpValPheGlyThrValIleGlyAla 160
 Db 421 ATGATCGCATTCGCTTCAGTTTCATGTCATCGTGGGTATTTGTACGGTAATAGGGCA 480
 Qy 161 PheSerGlySerGlyLeuLeuGlnGlyTyrProAlaValGluAlaAlaLeuGlyPheMet 180
 Db 481 TTCTCCGCGCAGCGCTTGTCTGCAAGGTTATCCCGCGTTGAAGCTGCATTAGGTTTATG 540
 Qy 181 LeuProAlaLeuPheMetSerPheLeuLeuAlaSerPheGlnArgLysGlnSerLeuCys 200
 Db 541 CTTCGCGCATTCTTTATGAGTTTCCTGCTCGCTCTTTCCAGCGCAACATCTCTTTC 600
 Qy 201 ValThrAlaAlaLeuValGlyAlaLeuAlaGlyValThrLeuPheSerIleProValAla 220
 Db 601 GTTACCGCAGCGTTAGTTGTGTCCTTGCAGCGTAAAGCTATTTTCTATTCCGCTGCC 660
 Qy 221 IleLeuAlaGlyIleValCysGlyCysLeuThrAlaLeuIleGlnAlaPheTrpGlnGly 240
 Db 661 ATTCGCGCAGCATTTGTCTGGCTGCTCACTCGTTAATCCAGCGCATCTTGGCAAGGA 720
 Qy 241 AlaProAspGluLeu 245
 Db 721 GCGCCCGATGAGCTA 735

RESULT 3

ID AD236194

XX AD236194 standard; DNA; 738 BP.

AC AD236194;

XX

DT 14-JUL-2005 (first entry)
 XX Escherichia coli b2862 encoding DNA SEQ ID NO:3.
 DE amino acid production; gene; ds.
 KW Escherichia coli.

OS
 XX Key Location/Qualifiers
 XX CDS 1..738
 FT /*tag= a
 FT /product= "b2862 protein"

XX EPI526181-A1.

XX 27-APR-2005.

XX 13-FEB-2002; 2004EP-00028877.

XX 13-FEB-2001; 2001RU-00103865.

PR 26-FEB-2001; 2001RU-00104998.

PR 26-FEB-2001; 2001RU-00104999.

PR 28-JUN-2001; 2001RU-00117632.

PR 28-JUN-2001; 2001RU-00117633.

PR 13-FEB-2002; 2002EP-00003335.

XX (AJIN) AJINOMOTO CO INC.

XX Tabolina EA, Rybak KV, Khourges EM, Voroshilova EB, Gueyatiner MM;

XX WPI; 2005-317157/33.

DR P-PSDB; AD236195.

XX New modified L-amino acid producing bacterium useful to enhance the
 PT production of L-amino acid by enhancing the activities of proteins in a
 PT cell of bacterium.

XX Disclosure; SEQ ID NO 3; 35pp; English.

XX The invention relates to an L-amino acid (A) producing bacterium (I)
 CC (belonging to the genus Escherichia), which is modified to enhance the
 CC production of (A) by enhancing the activities of proteins (G) or (H) in a
 CC cell of (I). Also described is a method for producing (A) comprising
 CC cultivating the bacterium in a culture medium and collecting the produced
 CC and accumulated L-amino acid. The modified bacterium has the ability to
 CC grow on a minimal medium containing L-amino acid or its analog in a
 CC minimal concentration and ability to grow faster on a medium containing L
 CC -amino acid or its analog than the unmodified strain or the wild type
 CC strain, or the parental strain of the bacterium. The present sequence
 CC encodes the E. coli b2862 protein which has L-amino acid excretion
 CC activity.

SQ Sequence 738 BP; 136 A; 192 C; 198 G; 212 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 8.45e-124 Length: 738
 Score: 1248.00 Matches: 245
 Percent Similarity: 100.0% Conservatives: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 100.0% Indels: 0
 DB: 14 Gaps: 0

US-10-073-293A-4 (1-245) x AD236194 (1-738)

Qy 1 MetGluSerProThrProGlnProAlaProGlySerAlaThrPheMetGluGlyCysLys 20

Db 1 ATGGAAGCCCTACTCCACACCTGCTCTGGTTCGGCGACCTTCATGGAAGGATGCAAA 60

Qy 21 AspSerLeuProIleValIleSerTyrIleProValAlaPheAlaPheGlyLeuAsnAla 40

Db 61 GACAGTTTACCGATTGTTATTAGTTATTTCCGGTGGCCCTTCGGTCTCGGATGCG 120

Qy 41 ThrArgLeuGlyPheSerProLeuGluSerValPhePheSerCysIleIleTyrAlaGly 60

```
Db 121 ACCGGTCGGGATTCCTCTCCCTCGAAGACGCTTTTCTCTCGATCATTTATGCGGC 180
Qy 61 AlaSerGlnPheValIleThrAlaMetLeuAlaGlySerSerLeuTrpIleAlaAla 80
Db 181 GCGAGCCAGTTCGTTCATACCGCGATGCTGCAGCCGGAGTAGTTTGTGGATTGCTGCA 240
Qy 81 LeuThrValMetAlaMetAspValArgHisValLeuTyrGlyProSerLeuArgSerArg 100
Db 241 CTGACCGGCATGGCAATGGATGTTCGCCATGTGTGTATGCCCCGTCACTCGTAGCCGT 300
Qy 101 IleIleGlnArgLeuGlnLysSerLysThrAlaLeuTrpAlaPheGlyLeuThrAspGlu 120
Db 301 ATTATTACGCGCTGCGAAATTCGAAACCCGCCCTGTGGGGCTTGGCTACCGATGAG 360
Qy 121 ValPheAlaAlaAlaThrAlaLysLeuValArgAsnAsnArgTTrpSerGluAsnTrp 140
Db 361 GTTTTGGCCGCGCAACCGCAAACTGGTACGCAATATCGCGCTGGAGCGAGAACTGG 420
Qy 141 MetIleGlyIleAlaPheSerSerTrpSerSerTrpValPheGlyThrValIleGlyAla 160
Db 421 ATGATCGGCATTCCTTCAGTTTCATGGTTCATGCTGTGGTATTTGGTACGGTAAATAGGGCA 480
Qy 161 PheSerGlySerGlyLeuLeuGlnGlyTyrProAlaValGluAlaAlaLeuGlyPheMet 180
Db 481 TTCTCCGCGACGGCTTCTCGAAGTTATCCGCCGTTGAAGCTGCATTAAGTTTATG 540
Qy 181 LeuProAlaLeuPheMetSerPheLeuLeuAlaSerPheGlnArgLysGlnSerLeuCys 200
Db 541 CTTCGGGCACCTTTATGAGTTTCTGCTCGCCTCTTTCCAGCGCAACAATCTCTTTCG 600
Qy 201 ValThrAlaAlaLeuValGlyAlaLeuAlaGlyValThrLeuPheSerIleProValAla 220
Db 601 GTTACCGCAGCGTTAGTTGGTGGCCCTTCGAGCGGTAACGCTATTCTTATTTCCCGTCGCC 660
Qy 221 IleLeuAlaGlyIleValCysGlyCysLeuThrAlaLeuIleGlnAlaPheTrpGlnGly 240
Db 661 ATTCTGGCAGCGATTCCTGTGGCTGCTCACTGCGTTAATCCAGGCATTTCTGGCAGGA 720
Qy 241 AlaProAspGluLeu 245
Db 721 GCGCCCGATGAGCTA 735
RESULT 4
ACH97221
ID ACH97221 standard; DNA; 756 BP.
XX AC ACH97221;
XX AC ACH97221;
DT 29-JUL-2004 (first entry)
DE Klebsiella pneumoniae polynucleotide seqid 3016.
XX KW Recombinant expression vector; transcription regulatory element;
XX KW Klebsiella pneumoniae protein; antibacterial; Vaccine; gene; ds.
XX OS Klebsiella pneumoniae.
XX FN US6610836-B1.
XX PD 26-AUG-2003.
XX PF 27-JAN-2000; 2000US-00489039.
XX PR 29-JAN-1999; 99US-0117747P.
XX PA (GENO-) GENOME THERAPEUTICS CORP.
XX PI Breton GL, Osborne M;
XX DR WPI; 2003-895346/82.
XX DR P-PSDB; ABO63670.
XX
```

```
PT New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for
PT preparing a vaccine composition against Klebsiella pneumoniae.
XX
PS Disclosure; SEQ ID NO 3016; 932pp; English.
XX
CC The invention describes a new isolated nucleic acid encoding a Klebsiella
CC pneumoniae polypeptide. Also described are: a recombinant expression
CC vector comprising the nucleic acid, operably linked to a transcription
CC regulatory element; and a cell comprising the recombinant expression
CC vector. The nucleic acid is useful for preparing a vaccine composition
CC against Klebsiella pneumoniae. This sequence encodes a Klebsiella
CC pneumoniae polypeptide of the invention
XX
SQ Sequence 756 BP; 110 A; 243 C; 229 G; 174 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.: 1.13e-98 Length: 756
Score: 1013.50 Matches: 200
Percent Similarity: 88.3% Conservative: 18
Best Local Similarity: 81.0% Mismatches: 26
Query Match: 81.2% Indels: 3
DB: 11 Gaps: 1
US-10-073-293A-4 (1-245) x ACH97221 (1-756)
Qy 1 MetGluSerProThrProGlnProAla-----ProGlySerAlaThrPheMetGlu 17
Db 10 ATGGAAACCCCGCCCTTAACTGTGCTTACTGAACGCGTCGCCACCGTCGGCGAA 69
Qy 18 GlyCysLysAspSerLeuProIleValIleSerTyrIleProValAlaPheAlaPheGly 37
Db 70 GCGCTCAAGACAGTCTCCCATCGTCATCAGCTACCTTCGGTGGCATTGCGCTTCGGA 129
Qy 38 LeuAsnAlaThrArgLeuGlyPheSerProLeuGluSerValPhePheSerCysIleIle 57
Db 130 CTCAATGTACCCGGCTGGGCTTCACCGCTGGAAGAGCTGTCTTCTCTCGCATATT 189
Qy 58 TyrAlaGlyAlaSerGlnPheValIleThrAlaMetLeuAlaAlaGlySerSerLeuTrp 77
Db 190 TACCGCGCGCCAGCCAGTTTGTGATCACCCCATGTTGGCCCGCGGAGTTGCTGTGG 249
Qy 78 IleAlaAlaLeuThrValMetAlaMetAspValArgHisValLeuTyrGlyProSerLeu 97
Db 250 GTCCGCGCCCTGACGCGTGTGATGCGATGTCGCCACGCTGCTATGTGGCCCTTCACTG 309
Qy 98 ArgSerArgIleIleGlnArgLeuGlnLysSerLysThrAlaLeuTrpAlaPheGlyLeu 117
Db 310 GCGAGCGCATTCGCTCAGCGCTGGATAAAAGAAAACCCGCCCTTGGGCCCTTGGCCTG 369
Qy 118 ThrAspGluValPheAlaAlaThrAlaLysLeuValArgAsnAsnArgTTrpSer 137
Db 370 ACCGATGAGGTTTTCGCCCGCGCTACCGCCAGGCTGGTTCGCGATAACCCCGCTGGAGC 429
Qy 138 GluAsnTrpMetIleGlyIleAlaPheSerSerTrpSerSerTrpValPheGlyThrVal 157
Db 430 GAAACCTGGATGCTCGGCTCGCTTTACCTTCGTGGGATCGTGGGTGGCGCACGCTG 489
Qy 158 IleGlyAlaPheSerGlySerGlyLeuLeuGlnGlyTyrProAlaValGluAlaAlaLeu 177
Db 490 GCCGCGCGCTGGTCCGCGCACGCGCTGCTGGTGGATTATCCGGCGGTGGAAAGCGCCCTC 549
Qy 178 GlyPheMetLeuProAlaLeuPheMetSerPheLeuLeuAlaSerPheGlnArgLysGln 197
Db 550 GGCTTTATGCTGCGCGCGCTGTTTATGAGCTTCTGTGCTGGCTCTTTTCCAGCTGACGAG 609
Qy 198 SerLeuCysValThrAlaAlaLeuValGlyAlaLeuAlaGlyValThrLeuPheSerIle 217
Db 610 TCACCTCGGTCACGCGCGCGCTGGCGGCGCGCTGGCGCGCAITCTCTTATTTTCCATT 669
Qy 218 ProValAlaIleLeuAlaGlyIleValCysGlyCysLeuThrAlaLeuIleGlnAlaPhe 237
Db 670 CCGCGCGCGATCTCTCGCGGAATTGCTGTGCGGGTGTCTGACGCGGCTGCTTTCAGGCCATG 729
```


WP	ACF67367_17	1700001	1810000
WP	ACF67367_18	1800001	1910000
WP	ACF67367_19	1900001	2010000
WP	ACF67367_20	2000001	2110000
WP	ACF67367_21	2100001	2210000
WP	ACF67367_22	2200001	2310000
WP	ACF67367_23	2300001	2410000
WP	ACF67367_24	2400001	2510000
WP	ACF67367_25	2500001	2610000
WP	ACF67367_26	2600001	2710000
WP	ACF67367_27	2700001	2810000
WP	ACF67367_28	2800001	2910000
WP	ACF67367_29	2900001	3010000
WP	ACF67367_30	3000001	3110000
WP	ACF67367_31	3100001	3210000
WP	ACF67367_32	3200001	3310000
WP	ACF67367_33	3300001	3410000
WP	ACF67367_34	3400001	3510000
WP	ACF67367_35	3500001	3610000
WP	ACF67367_36	3600001	3710000
WP	ACF67367_37	3700001	3810000
WP	ACF67367_38	3800001	3910000
WP	ACF67367_39	3900001	4010000
WP	ACF67367_40	4000001	4110000
WP	ACF67367_41	4100001	4210000
WP	ACF67367_42	4200001	4310000
WP	ACF67367_43	4300001	4410000
WP	ACF67367_44	4400001	4510000
WP	ACF67367_45	4500001	4610000
WP	ACF67367_46	4600001	4710000
WP	ACF67367_47	4700001	4810000
WP	ACF67367_48	4800001	4910000
WP	ACF67367_49	4900001	5010000
WP	ACF67367_50	5000001	5110000
WP	ACF67367_51	5100001	5210000
WP	ACF67367_52	5200001	5310000
WP	ACF67367_53	5300001	5410000
WP	ACF67367_54	5400001	5510000
WP	ACF67367_55	5500001	5610000
WP	ACF67367_56	5600001	5648894
Alignment Scores:			
Pred. No.:	3,16e-77	Length:	110000
Score:	841.50	Matches:	160
Percent Similarity:	82.1%	Conservative:	32
Best Local Similarity:	68.4%	Mismatches:	41
Query Match:	67.4%	Indels:	1
DB:	10	Gaps:	1
US-10-073-293A-4 (1-245) x ACF67367_20 (1-110000)			
Qy	3 SerProThrProGlnProAlaProGly---SerAlaThrPheMetGluGlyCysLysAsp	21	
Db	22141 TCTGATACATCATCCCCCTTAACGAGTAAATAATCTCTTTACTGAAGGAATAATAGAT	22200	
Qy	22 SerLeuProIleValIleSerTyrIleProValAlaPheAlaPheGlyLeuAsnAlaThr	41	
Db	22201 AGTTTACCATTGTTATCCGGTTATATCCCGTCGCTTTGCTTCAATGCCGTC	22260	
Qy	42 ArgLeuGlyPheSerProLeuGluSerValPhePheSerCysIleIleTyrAlaGlyAla	61	
Db	22261 AAACCTTGCTCAACCCCAATGAAGCCATTTCTTTTCATGCATCATACGCCGTGCA	22320	
Qy	62 SerGlnPheValIleThrAlaMetLeuAlaAlaGlySerSerLeuTrpIleAlaAlaLeu	81	
Db	22331 AGCCAGTTTCGTCATCACAGCTTTACTGAGTGGGGGACATCATATTATGATTTCTGCCCTA	22380	
Qy	82 ThrValMetAlaMetAspValArgHisValLeuTyrGlyProSerLeuArgSerArgile	101	
Db	22381 ACAATTATGGCAATGGATGTCGCGCATATTCTTTATGTGTCATCTTTTAAGCACCGTATC	22440	
Qy	102 IleGlnArgLeuGlnLysSerLysThrAlaLeuTrpAlaPheGlyLeuThrAspGluVal	121	

Db	22441 AAAGATAAGCTAACGGAGAAAAAACCGTTATCTGGGCTTTCGGCCTGCACAGATGAAGTT	22500	
Qy	122 PheAlaAlaThrAlaLysLeuValArgAsnAsnArgTTrpSerGluAsnTrpMet	141	
Db	22501 TTTGCCCGCGGACTGCATAAATCATTAATAAACCCCGAGCTGGAGTGAANAATGGATG	22560	
Qy	142 IleGlyIleAlaPheSerSerTrpValPheGlyThrValIleGlyAlaPhe	161	
Db	22561 GTTCTATTGCAATCTGTTCTTGGCTGGCGCGCAGGTACCGACGCGGTGCAITTT	22620	
Qy	162 SerGlySerGlyLeuLeuGlnGlyTyrProAlaValGluAlaAlaLeuGlyPheMetLeu	181	
Db	22621 CTTGGTAACGGTTATTTCGAATCTATCCGCTATAGAAGCTGCATATTTTCATGTTA	22680	
Qy	182 ProAlaLeuPheMetSerPheLeuAlaSerPheGlnArgLysGlnSerLeuCysVal	201	
Db	22681 CCAGCACTATTCTCAGTTTCTTCTGCTTCTGTAGAAAACAAATAGTATTGTGTT	22740	
Qy	202 ThrAlaAlaLeuValGlyAlaLeuAlaGlyValThrLeuPheSerIleProValAlaIle	221	
Db	22741 GCAACCGCACTAACCGGAGCACCTTTTAGGGATTACATTTTCTCAATTCAGTTGCTATT	22800	
Qy	222 LeuAlaGlyIleValCysGlyCysLeuThrAlaLeuIleGln	235	
Db	22801 CTGGCAGGTATTGTGGTGGTTGTATCGCGCACTGTTACAA	22842	
RESULT 7			
ACF65382			
ID	ACF65382 standard; DNA; 243072 BP.		
XX	ACF65382;		
DT	20-NOV-2003 (first entry)		
XX	Photorhabdus luminescens nucleotide sequence #35.		
XX	Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;		
XX	detection; food; gene expression; plant; animal; microorganism; toxin;		
XX	antibiotic; biopesticide; virulence factor; disease model; plague;		
XX	whooping cough; gene; ds.		
XX	Photorhabdus luminescens.		
XX	WO200294867-A2.		
XX	28-NOV-2002.		
XX	07-FEB-2002; 2002WO-IB003040.		
XX	07-FEB-2001; 2001FR-00001659.		
XX	(INSP) INST PASTEUR.		
XX	(CNRS) CNRS CENT NAT RECH SCI.		
PI	Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;		
PI	Buchrieser C;		
XX	WPI; 2003-148459/14.		
XX	Genomic sequence of Photorhabdus luminescens and encoded polypeptides,		
XX	useful e.g. as therapeutic antimicrobials and agricultural pesticides.		
PS	Claim 1; SEQ ID NO 35; 1205pp; French.		
XX	The invention relates to the isolation of genes and their encoded		
XX	proteins from Photorhabdus luminescens. The isolated sequences are		
XX	sources of probes and primers for detecting the genome of P. luminescens		
XX	and related species; to study polymorphisms; for gene analysis and for		
XX	detection/amplification of the genes. Antibodies (Ab) raised against the		
XX	polypeptides encoded by the genes are used for detection/identification		
XX	of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that		
XX	carry a gene-containing vector are used to select compounds that		
XX	modulate, regulate, induce or inhibit expression of the genes in plants,		

CC animals or microorganisms other than *P. luminescens* and are able to alter
 CC response or sensitivity to toxins and antibiotics produced by *P.*
 CC *luminescens*. Cells transformed to express the genes are useful for
 CC recombinant production of the proteins, particularly toxins and
 CC antibacterial useful as insecticides, bactericides and fungicides. The
 CC genes, proteins, vectors containing the genes and Ab are also useful
 CC therapeutically (to treat microbial infection by bacteria or fungi that
 CC are sensitive to *P. luminescens*-encoded toxins or antibiotics) and as
 CC biopesticides. Other uses of the genes and the proteins are as virulence
 CC factors and for identifying targets of human diseases for which *P.*
 CC *luminescens* is a model (particularly plaque and whooping cough). This
 CC sequence represents one of the isolated *P. luminescens* genes
 XX
 SQ Sequence 243072 BP; 70273 A; 57885 C; 49427 G; 65487 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 9,59e-77 Length: 243072
 Score: 841.50 Matches: 160
 Percent Similarity: 82.1% Conservatives: 32
 Best Local Similarity: 68.4% Mismatches: 41
 Query Match: 67.4% Indels: 1
 DB: 10 Gaps: 1

US-10-073-293A-4 (1-245) x ACF65382 (1-243072)

Qy 3 SerProThrProGlnProAlaProGly---SerAlaThrPheMetGluGlyCysLysAsp 21
 Db 74883 TCTGATACATCATCCCCCTTAACAGTAAATAAATCTCTTTTACTGAAGGAATAATAGAT 74942
 Qy 22 SerLeuProIleValIleSerTyrlleProValAlaPheAlaPheGlyLeuAsnAlaThr 41
 Db 74943 AGTTTACCATTGTTATCGGTTATATCCCGTCGCTTGTCTTGTCTCAATGCCGTC 75002
 Qy 42 ArgLeuGlyPheSerProLeuGlySerValPhePheSerCysIleIleTyrlleAlaGlyVala 61
 Db 75003 AACTTGGCTTCAACCCCAATGGAAGCCATTTCTTTTCATGCATCATCTAGCCGGTGCA 75062
 Qy 62 SerGlnPheValIleThrAlaMetLeuAlaAlaGlySerSerLeuThrTrpIleAlaAlaLeu 81
 Db 75063 AGCCAGTTCGTCATCACAGCTTTACTGAGTCGGGGACATCATTTATGGATTCTGGCCTA 75122
 Qy 82 ThrValMetAlaMetAspValArgHisValLeuTyrlleGlyProSerLeuArgSerArgIle 101
 Db 75123 ACAATTATGCAATGGATGTCCGCATATCTTTTATGTGTCATCTTTAAGGCACCGTATC 75182
 Qy 102 IleGlnArgLeuGlnLysSerLysThrAlaLeuThrAlaPheGlyLeuThrAspGluVal 121
 Db 75183 AAGATAGCTAACCGGAGAAAAACCGTTATCTGGGCTTTTCGGCCTGACAGATGAAGTT 75242
 Qy 122 PheAlaAlaThrAlaLysLeuValArgAsnAsnArgArgTrpSerGluAsnTrpMet 141
 Db 75243 TTTGCCGCGGACTGCAAACTCATTAATAAACCCAGCGAGCTGGAGTGAACCTGGATG 75302
 Qy 142 IleGlyIleAlaPheSerSerTrpSerTrpValPheGlyThrValIleGlyAlaPhe 161
 Db 75303 GTTGCTATTGCAATCTGTTCTTGGCTGCGCTGGCGCGAGTACCGCAGCGCGTGCATTT 75362
 Qy 162 SerGlySerGlyLeuLeuGlnGlyTyrlleProValAlaGluAlaAlaLeuGlyPheMetLeu 181
 Db 75363 CTTCGTAAACGGTTATTGGAAATCTTATCCCGCTATAGAAGCTGCCATGATTTTCATGTTA 75422
 Qy 182 ProAlaLeuPheMetSerPheLeuAlaSerPheGlnArgLysGlnSerLeuCysVal 201
 Db 75423 CCACACTATTCTCAGTTTCTTCTGCTTCTTGTAGAACAAATAATAGTTATTGTTT 75482
 Qy 202 ThrAlaAlaLeuValGlyAlaLeuAlaGlyValThrLeuPheSerIleProValAlaIle 221
 Db 75483 GCAACCGCACTAACCGGAGCACTTTTAGGGATACATTTTCTCAATTCAGTTGCTATT 75542
 Qy 222 LeuAlaGlyIleValCysGlyCysLeuThrAlaLeuIleGln 235
 Db 75543 CTGGCAGTATTGTGGTGGTGTATCGCGCACTGTTTACAA 75584

RESULT 8

ADA31112
 ID ADA31112 standard; DNA; 801 BP.
 XX
 AC ADA31112;
 XX
 DT 20-NOV-2003 (first entry)
 XX
 DE DNA encoding Acinetobacter baumannii protein #2399.
 XX
 KW ds; gene; Acinetobacter baumannii; bacterial disease; antibacterial;
 KW vaccine; plant biocontrol agent.
 XX
 OS Acinetobacter baumannii.
 XX
 FN US6562958-B1.
 XX
 PD 13-MAY-2003.
 XX
 PF 04-JUN-1999; 99US-00328352.
 XX
 PR 09-JUN-1998; 98US-0088701P.
 XX
 PA (GENO-) GENOME THERAPEUTICS CORP.
 XX
 PI Breton G, Bush D;
 XX
 DR WPI; 2003-576092/54.
 DR P-PSDB; ADA35238.
 XX
 PT New Acinetobacter baumannii proteins and nucleic acids, useful as reagents
 PT for diagnosing a bacterial disease, as components of antibacterial
 PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for
 PT plants.
 XX
 PS Example; SEQ ID NO 2399; 328pp; English.
 XX
 CC The invention relates to isolated Acinetobacter baumannii nucleic acids.
 CC The A. baumannii nucleic acids and polypeptides are useful as reagents
 CC for diagnosing a bacterial disease, as components of antibacterial
 CC vaccines, as targets for antibacterial drugs, to detect the presence of
 CC A. baumannii and other Acinetobacter species in a sample, in screening
 CC compounds for the ability to interfere with the A. baumannii life cycle
 CC or to inhibit A. baumannii infection, and as biocontrol agents for
 CC plants. The present sequence represents DNA encoding an A. baumannii
 CC protein.
 XX
 SQ Sequence 801 BP; 196 A; 147 C; 165 G; 293 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 1,67e-53 Length: 801
 Score: 592.00 Matches: 111
 Percent Similarity: 71.0% Conservatives: 53
 Best Local Similarity: 48.1% Mismatches: 67
 Query Match: 47.4% Indels: 0
 DB: 9 Gaps: 0

US-10-073-293A-4 (1-245) x ADA31112 (1-801)

Qy 7 GlnProAlaProGlySerAlaThrPheMetGluGlyCysLysAspSerLeuProIleVal 26
 Db 76 CAGATGGCTGCTCAACAGCAACTTTTGGCAAGGTGCCAAGATAGTCAAGCCATCGTT 135
 Qy 27 IleSerTyrlleProValAlaPheAlaPheGlyLeuAsnAlaThrArgLeuGlyPheSer 46
 Db 136 CTCACCTACTTACAGATTCTTTTGGCTTTTGGCGTTCTGCATCTCAGTTGGTTTACA 195
 Qy 47 ProLeuGluSerValPhePheSerCysIleIleTyrlleAlaGlyAlaSerGlnPheValIle 66
 Db 196 CCATGGGAAGCAATTTTCTTTTCTTGTTCATGTATGAGTGCAGGTGCAAGCCAAATTTTAGTG 255
 Qy 67 ThrAlaMetLeuAlaAlaGlySerSerLeuTrpIleAlaLeuThrValMetAlaMet 86

```
Db      256  GTCCGACATTTAGCAAGTGGTTCTCTATATAGTTTAACCGCTTTAAACAGTCATTGCACTT 315
Qy      87  AspValArgHisValLeuTyGlyProSerLeuArgSerArgIleIleGlnArgLeuGln 106
Db      316  GATATACACATGACTTTATGGGCTCTCTTTATTAATTTCCACATAAATTAAT 375
Qy      107  LysSerLysThrAlaLeuTyrAlaPheGlyLeuThrAspGluValPheAlaAlaThr 126
Db      376  TTAAGAAGACAGCGTTTGGGCTGGGTTTACGGACGAAGTTTTTGCAGTGGTATG 435
Qy      127  AlalysLeuValArgAsnAsnArgTyrSerGluAsnTyrMetIleGlyIleAlaPhe 146
Db      436  ATCCAACATTTACAGCGCAGACAACAATGGTCCGAGTCTTGGATGTTGGGCTTAAGTCTA 495
Qy      147  SerSerTyrSerSerTyrValPheGlyThrValIleGlyAlaPheSerGlySerGlyLeu 166
Db      496  TTTAGCTGGATGTCCTTGGGCAAGTGGTTCATTAATTAGCGGGTATTATTGCGAGATCAAGTC 555
Qy      167  LeuGlnGlyTyrProAlaValGluAlaAlaLeuGlyPheMetLeuProAlaLeuPheMet 186
Db      556  GCACACCTTCAAGTTTTTACAGCACCTTTAGATTCTTATTACCTGCTTTATTTTAA 615
Qy      187  SerPheLeuLeuAlaSerPheGlnArgLysGlnSerLeuCysValThrAlaAlaLeuVal 206
Db      616  AGTTTTTTATTAGCGGCATTTGAACGAAGCATAGCCTTGTAGTTGTGTATCGTTAGTGT 675
Qy      207  GlyAlaLeuAlaGlyValThrLeuPheSerIleProValAlaIleLeuAlaGlyIleVal 226
Db      676  GTGTCAGCTTTGGCATGTTATTGGATTAACTATACGCGGCTATCTTTATTGGCATTCCT 735
Qy      227  CysGlyCysLeuThrAlaLeuIleGlnAlaPhe 237
Db      736  TCAGGCATTTAGCAGGCTATTAAATATTAT 768

RESULT 9
ID      AAL35397 standard; cDNA; 1589 BP.
XX
AC      AAL35397;
XX
DT      08-JAN-2002 (first entry)
XX
DE      Human musculoskeletal system related polynucleotide SEQ ID NO 739.
XX
KW      Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW      antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW      vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW      cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW      neurological disease; infection; human; secreted protein;
KW      musculoskeletal system; ss.
XX
OS      Homo sapiens.
XX
PN      WO200155367-A1.
XX
PD      02-AUG-2001.
XX
PF      17-JAN-2001; 2001WO-US001338.
XX
PR      31-JAN-2000; 2000US-0179065P.
PR      04-FEB-2000; 2000US-0180628P.
PR      24-FEB-2000; 2000US-0184664P.
PR      02-MAR-2000; 2000US-0186350P.
PR      16-MAR-2000; 2000US-0189874P.
PR      17-MAR-2000; 2000US-0190076P.
PR      18-APR-2000; 2000US-0198123P.
PR      19-MAY-2000; 2000US-0205515P.
PR      07-JUN-2000; 2000US-0209467P.
PR      28-JUN-2000; 2000US-0214886P.
PR      30-JUN-2000; 2000US-0215135P.
PR      07-JUL-2000; 2000US-0216647P.
PR      07-JUL-2000; 2000US-0216880P.
PR      11-JUL-2000; 2000US-0217487P.
PR      11-JUL-2000; 2000US-0217496P.
PR      14-JUL-2000; 2000US-0218290P.
PR      26-JUL-2000; 2000US-0220963P.
PR      26-JUL-2000; 2000US-0220964P.
PR      14-AUG-2000; 2000US-0224518P.
PR      14-AUG-2000; 2000US-0224519P.
PR      14-AUG-2000; 2000US-0225213P.
PR      14-AUG-2000; 2000US-0225214P.
PR      14-AUG-2000; 2000US-0225266P.
PR      14-AUG-2000; 2000US-0225267P.
PR      14-AUG-2000; 2000US-0225268P.
PR      14-AUG-2000; 2000US-0225270P.
PR      14-AUG-2000; 2000US-0225447P.
PR      14-AUG-2000; 2000US-0225757P.
PR      14-AUG-2000; 2000US-0225758P.
PR      14-AUG-2000; 2000US-0225759P.
PR      18-AUG-2000; 2000US-0226279P.
PR      22-AUG-2000; 2000US-0226681P.
PR      22-AUG-2000; 2000US-0226868P.
PR      22-AUG-2000; 2000US-0227182P.
PR      23-AUG-2000; 2000US-0227009P.
PR      30-AUG-2000; 2000US-0228924P.
PR      01-SEP-2000; 2000US-0229287P.
PR      01-SEP-2000; 2000US-0229343P.
PR      01-SEP-2000; 2000US-0229344P.
PR      01-SEP-2000; 2000US-0229345P.
PR      05-SEP-2000; 2000US-0229509P.
PR      05-SEP-2000; 2000US-0229513P.
PR      06-SEP-2000; 2000US-0230437P.
PR      06-SEP-2000; 2000US-0230438P.
PR      08-SEP-2000; 2000US-0231242P.
PR      08-SEP-2000; 2000US-0231243P.
PR      08-SEP-2000; 2000US-0231244P.
PR      08-SEP-2000; 2000US-0231413P.
PR      08-SEP-2000; 2000US-0231414P.
PR      08-SEP-2000; 2000US-0232080P.
PR      12-SEP-2000; 2000US-0231968P.
PR      14-SEP-2000; 2000US-0232397P.
PR      14-SEP-2000; 2000US-0232398P.
PR      14-SEP-2000; 2000US-0232399P.
PR      14-SEP-2000; 2000US-0232400P.
PR      14-SEP-2000; 2000US-0232401P.
PR      14-SEP-2000; 2000US-0233063P.
PR      14-SEP-2000; 2000US-0233084P.
PR      21-SEP-2000; 2000US-0233085P.
PR      21-SEP-2000; 2000US-0234223P.
PR      21-SEP-2000; 2000US-0234274P.
PR      25-SEP-2000; 2000US-0234997P.
PR      25-SEP-2000; 2000US-0234998P.
PR      26-SEP-2000; 2000US-0235484P.
PR      27-SEP-2000; 2000US-0235834P.
PR      27-SEP-2000; 2000US-0235836P.
PR      29-SEP-2000; 2000US-0236327P.
PR      29-SEP-2000; 2000US-0236367P.
PR      29-SEP-2000; 2000US-0236368P.
PR      29-SEP-2000; 2000US-0236369P.
PR      02-OCT-2000; 2000US-0236802P.
PR      02-OCT-2000; 2000US-0237037P.
PR      02-OCT-2000; 2000US-0237038P.
PR      02-OCT-2000; 2000US-0237039P.
PR      02-OCT-2000; 2000US-0237040P.
PR      13-OCT-2000; 2000US-0239935P.
PR      13-OCT-2000; 2000US-0239937P.
PR      20-OCT-2000; 2000US-0240960P.
PR      20-OCT-2000; 2000US-0241221P.
PR      20-OCT-2000; 2000US-0241785P.
PR      20-OCT-2000; 2000US-0241786P.
PR      20-OCT-2000; 2000US-0241787P.
PR      20-OCT-2000; 2000US-0241808P.
PR      20-OCT-2000; 2000US-0241809P.
PR      20-OCT-2000; 2000US-0241826P.
```

PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249246P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 11-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-02559678P.
 (HUMA-) HUMAN GENOME SCI INC.

XX PA
 XX PI
 XX Rosen CA, Barash SC, Ruben SM;
 XX WPI; 2001-451937/48.
 DR P-PSDB; ABB03815.

XX Isolated polypeptide for treating, preventing and/or prognosing
 PT disorders related to the musculoskeletal system including musculoskeletal
 PT cancers and also for testing and detection e.g. diagnosis.

XX Claim 1; SEQ ID NO 739; 781pp + Sequence Listing; English.

XX The invention relates to novel genes (AAL34669-AAL37666) and proteins
 CC (ABB03087-ABB04109) associated with the musculoskeletal system useful for
 CC preventing, treating or ameliorating medical conditions e.g. by protein
 CC or gene therapy. The genes are isolated from a range of human tissues
 CC disclosed in the specification. The nucleic acids, proteins, antibodies
 CC and (ant)agonists are useful in the diagnosis, treatment and prevention
 CC of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the
 CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
 CC lung, or urogenital; (b) immune disorders e.g. Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,

CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as
 CC myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g.
 CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
 CC bacterial, fungal and parasitic infections. Note: The sequence data for
 CC this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX

SQ Sequence 1589 BP; 307 A; 412 C; 414 G; 456 T; 0 U; 0 Other;

Alignment Scores:

Pred. NO.: 2.61e-35 Length: 1589
 Score: 426.00 Matches: 84
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 34.1% Indels: 0
 DB: 4 Gaps: 0

US-10-073-293A-4 (1-245) x AAL35397 (1-1589)

Qy 1 MetGluSerProThrProGlnProAlaProGlySerAlaThrPheMetGluGlyCysIys 20
 |||||
 Db 1338 ATGGAAGCCCTACTCCACGCCCTGCTCGTTGGCGACCTTCATGGAAGGATGCANA 1397
 |||||
 Qy 21 AspSerLeuProIleValIleSerTyrIleProValAlaPheAlaPheGlyLeuAsnAla 40
 |||||
 Db 1398 GACAGTTTACCGATTGTTATTAGTTATATCCGGTGGCTTGGTTCGCTCTCAATGCG 1457
 |||||
 Qy 41 ThrArgLeuGlyPheSerProGluSerProLeuGluSerValPhePheSerCysIleIleTyrAlaGly 60
 |||||
 Db 1458 ACCGTCCTGGGATTCTCTCTCGAAGCGTTTTTTTCTCTGCATCATTTATGCGAGC 1517
 |||||
 Qy 61 AlaserGlnPheValIleThrAlaMetLeuAlaGlySerSerLeuTrioIleAlaAla 80
 |||||
 Db 1518 GCGAGCCAGTTCGTTCATTACCGCATCTGCGAGCGGAGTAGTTTGTGATTGCTGCA 1577
 |||||
 Qy 81 LeuThrValMet 84
 |||||
 Db 1578 CTGACCGTCATG 1589

RESULT 10

AAI62810
 ID AAI62810 standard; cDNA; 1589 BP.
 AC AAI62810;
 XX
 XX 22-OCT-2001 (first entry)
 XX Human cDNA SEQ ID NO 69.

XX Human; nontropic; neuroprotective; cytostatic; dermatological; virucide;
 KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
 KW antiparkinsonian; antiskilling; antianaemic; antiarthritic; cancer;
 KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
 KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
 KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ss.
 XX Homo sapiens.
 XX WO200155449-A1.
 XX PD 02-AUG-2001.
 XX PF 17-JAN-2001; 2001WO-US0001346.
 XX PR 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 19-MAY-2000; 2000US-0205515P.
 PR 07-JUL-2000; 2000US-0216880P.
 PR 14-JUL-2000; 2000US-0218290P.
 PR 14-AUG-2000; 2000US-0225447P.

```
PR 01-SEP-2000; 2000US-0229343P.
PR 06-SEP-2000; 2000US-0230437P.
PR 08-SEP-2000; 2000US-0231243P.
PR 25-SEP-2000; 2000US-0234997P.
PR 29-SEP-2000; 2000US-0236367P.
PR 13-OCT-2000; 2000US-0239937P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246528P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249265P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-476225/51.
XX P-PSDB; AAM42405.
XX
XX Novel plasma membrane associated proteins useful for diagnosing,
XX treating, preventing and/or prognosing disorders related to the proteins,
XX including cancer, immune response and neuronal disorders.
XX
XX Claim 1; SEQ ID NO 69; 532pp + Sequence Listing; English.
XX
XX The invention relates to novel genes (AAI62752-AAI62961) and proteins
XX (AAM42347-AAM42415) useful for preventing, treating or ameliorating
XX medical conditions e.g. by protein or gene therapy. The genes are
XX isolated from a range of human tissues disclosed in the specification.
XX The nucleic acids, proteins, antibodies and (ant)agonists are useful in
XX the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
XX ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
XX breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
XX disorders e.g. Addison's disease, allergies, autoimmune haemolytic
XX anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
XX multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
XX cardiovascular disorders such as myocardial ischaemias; (d) wound healing
XX ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
XX infectious diseases such as viral, bacterial, fungal and parasitic
XX infections. Note: The sequence data for this patent did not form part of
XX the printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 1589 BP; 307 A; 412 C; 414 G; 456 T; 0 U; 0 Other;
XX
XX Alignment Scores:
XX Pred. No.: 2,61e-35 Length: 1589
XX Score: 426.00 Matches: 84
XX Percent Similarity: 100.0% Conservative: 0
XX Best Local Similarity: 100.0% Mismatches: 0
XX Query Match: 34.1% Indels: 0
XX DB: 4 Gaps: 0
XX
XX US-10-073-293A-4 (1-245) x AAI62810 (1-1589)
XX
XX QY 1 MetGluSerProThrProGlnProAlaProGlySerAlaThrPheMetGluCysLys 20
XX
XX Db 1338 ATGGAAGCCCTACTCCAGCCTCTCTCGTTCGGCGACCTTTCATGGAAGGATGCAAA 1397
XX
XX QY 21 AspSerLeuProIleValIleSerTyrlleProValAlaPheAlaPheGlyLeuAsnAla 40
```

```
Db 1398 GACAGTTTACCGATTGTTATTAGTTATATTCGGTCCGCTTTCGGTTCGATGCG 1457
QY 41 ThrArgLeuGlyPheSerProLeuGluSerValPhePheSerCysIleIleTyrAlaGly 60
Db 1458 ACCGCTCTGGGATTCTCTCCTCTCGAAAGCGTTTTTTTCTCTCCTCATCATTTATGCGAGGC 1517
QY 61 AlaSerGlnPheValIleThrAlaMetLeuAlaAlaGlySerSerLeuTyrIleAlaAla 80
Db 1518 GCGAGCCAGTTCGTCATTACCGCATGCTGCAGCCGCGAGTAGTTTGTGGATTGCTGCA 1577
QY 81 LeuThrValMet 84
Db 1578 CTGACCGTCATG 1589
XX
XX RESULT 11
XX ABK43465
XX ID ABK43465 standard; cDNA; 1589 BP.
XX
XX AC ABK43465;
XX
XX DT 05-JUN-2002 (first entry)
XX
XX DE DNA encoding novel central nervous system protein #45.
XX
XX KW Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;
XX hyperproliferative disorder; neoplasm; cardiovascular disorder;
XX cardiac arrest; cerebrovascular disorder; ischaemia; angiogenesis;
XX nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;
XX acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;
XX adenocarcinoma; reproductive system disorder; testicular feminisation;
XX endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;
XX respiratory disorder; renal disorder; kidney failure; blood disorder;
XX myocardial infarction; wound healing; cell proliferation; skin aging;
XX food additive; food preservative; gene therapy; gene; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200155318-A2.
XX
XX PD 02-AUG-2001.
XX
XX PF 17-JAN-2001; 2001WO-US001332.
XX
XX PR 31-JAN-2000; 2000US-0179065P.
XX PR 04-FEB-2000; 2000US-0180628P.
XX PR 24-FEB-2000; 2000US-0184664P.
XX PR 02-MAR-2000; 2000US-0186350P.
XX PR 16-MAR-2000; 2000US-0189874P.
XX PR 17-MAR-2000; 2000US-0190076P.
XX PR 18-APR-2000; 2000US-0198123P.
XX PR 19-MAY-2000; 2000US-0205515P.
XX PR 28-JUN-2000; 2000US-0209467P.
XX PR 30-JUN-2000; 2000US-0214886P.
XX PR 07-JUL-2000; 2000US-0215135P.
XX PR 07-JUL-2000; 2000US-0216647P.
XX PR 07-JUL-2000; 2000US-0216880P.
XX PR 11-JUL-2000; 2000US-0217487P.
XX PR 11-JUL-2000; 2000US-0217496P.
XX PR 26-JUL-2000; 2000US-0218290P.
XX PR 26-JUL-2000; 2000US-0220963P.
XX PR 14-AUG-2000; 2000US-0220964P.
XX PR 14-AUG-2000; 2000US-0224518P.
XX PR 14-AUG-2000; 2000US-0224519P.
XX PR 14-AUG-2000; 2000US-0225213P.
XX PR 14-AUG-2000; 2000US-0225214P.
XX PR 14-AUG-2000; 2000US-0225266P.
XX PR 14-AUG-2000; 2000US-0225267P.
XX PR 14-AUG-2000; 2000US-0225268P.
XX PR 14-AUG-2000; 2000US-0225270P.
XX PR 14-AUG-2000; 2000US-0225447P.
XX PR 14-AUG-2000; 2000US-0225577P.
XX PR 14-AUG-2000; 2000US-0225758P.
```


PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0233223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.

PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 01-DEC-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 06-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 11-DEC-2000; 2000US-0251990P.
PR 05-JAN-2001; 2001US-0254097P.
PR 05-JAN-2001; 2001US-0254097P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-581633/65.
DR P-PSDB; AAU87135.
DR
XX
PT New isolated nucleic acid encoding a protein for diagnosing, preventing,
PT treating or ameliorating medical conditions and used as food additives or
PT preservatives.
XX
PS Claim 1; SEQ ID NO 55; 837pp; English.
XX
CC The invention describes an isolated nucleic acid molecule (I) encoding a
CC novel central nervous system protein. (I) and polypeptides (II) encoded
CC by (I), are used to treat a medical conditions and in diagnosis of a
CC pathological condition. Disorders which are diagnosed or treated include
CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease and
CC amyotrophic lateral sclerosis, infections caused by bacteria, viruses
CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,
CC adenocarcinomas and irritable bowel syndrome, reproductive system
CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.
CC leukemia, disorders involving neovascularisation e.g. malignancies,
CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
CC acute kidney failure and blood related disorders e.g. myocardial
CC infarction. The polypeptides can also be used to aid wound healing and
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities, fat content, lipid, protein,

Alignment Scores:

Pred. No.:	2.61e-35	Length:	1589
Score:	426.00	Matches:	84
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	34.1%	Indels:	0
DB:	4	Gaps:	0
US-10-073-293A-4 (1-245) x ABK43465 (1-1589)			
Qy	1	MetGluSerProThrProGlnProAlaProGlySerAlaThrPheMetGluGlyCysLys	20
Db	1338	ATGGAAGCCCTACTCCACAGCTCTCTCGTTCGGCGACCTTCATGGAAGGATGCAAA	1397
Qy	21	AspSerLeuProIleValIleSerTyrlleProValAlaPheAlaPheGlyLeuAsnAla	40
Db	1398	GACAGTTTACCGATTGTTATTAGTTATATTCGGTGGCTTTTGGCTTCGATGCG	1457
Qy	41	ThrArgLeuGlyPheSerProLeuGluSerValPhePheSerCysIleIleTyxAlaGly	60
Db	1458	ACCGTCTGGGATCTCTCTCTCGAAGCGTTTTTCTCTGTCATCATTTATGCAGGC	1517
Qy	61	AlaSerGlnPheValIleThrAlaMetLeuAlaAlaGlySerSerLeuTrpIleAlaAla	80
Db	1518	GCGAGCCAGTTCGTCTATACCGCGATGCTGCAGCCGGAGTAGTTTGTGGATTGCTGCA	1577
Qy	81	LeuThrValMet	84
Db	1578	CTGACCGTCATG	1589
RESULT 12			
ID	ADM19434	standard; cDNA; 1589 BP.	
XX	AC	ADM19434;	
XX	DT	20-MAY-2004 (first entry)	
XX	DE	Novel human channel/transporter gene #231.	
KW	ds; gene; immunosuppressive; antiarthritic; antirheumatic;		
KW	antiproliferative; cytostatic; cardiatic; vasotropic; cerebroprotective;		
KW	nocrotropic; neuroprotective; antibacterial; virucide; fungicide;		
KW	ophthalmological; gene therapy; channel/transporter protein;		
KW	rheumatoid arthritis; neoplasia; cardiac arrest; cerebrovascular disorder;		
KW	cerebral ischemia; angiogenesis; nervous system disorder;		
KW	Alzheimer's disease; ocular disorder; corneal infection; wound healing;		
KW	epithelial cell proliferation; skin aging; sunburn; transplantation;		
KW	chemotaxis; food additive.		
OS	Homo sapiens.		
XX	XX	WO200154472-A2.	
XX	PD	02-AUG-2001.	
XX	PF	17-JAN-2001; 2001WO-US001307.	
XX	PR	31-JAN-2000; 2000US-0179065P.	
PR	04-FEB-2000;	2000US-0180628P.	
PR	24-FEB-2000;	2000US-0184664P.	
PR	02-MAR-2000;	2000US-0186350P.	
PR	16-MAR-2000;	2000US-0189874P.	
PR	17-MAR-2000;	2000US-0190076P.	
PR	18-APR-2000;	2000US-0198123P.	
PR	19-MAY-2000;	2000US-0205515P.	
PR	07-JUN-2000;	2000US-0209467P.	
PR	28-JUN-2000;	2000US-0214886P.	
PR	30-JUN-2000;	2000US-0215135P.	
PR	07-JUL-2000;	2000US-0216647P.	
PR	07-JUL-2000;	2000US-0216880P.	
PR	11-JUL-2000;	2000US-0217487P.	
PR	11-JUL-2000;	2000US-0217496P.	
PR	14-JUL-2000;	2000US-0218290P.	
PR	26-JUL-2000;	2000US-0220963P.	
PR	26-JUL-2000;	2000US-0220964P.	
PR	14-AUG-2000;	2000US-0224518P.	
PR	14-AUG-2000;	2000US-0224519P.	
PR	14-AUG-2000;	2000US-0225213P.	
PR	14-AUG-2000;	2000US-0225214P.	
PR	14-AUG-2000;	2000US-0225266P.	
PR	14-AUG-2000;	2000US-0225267P.	
PR	14-AUG-2000;	2000US-0225288P.	
PR	14-AUG-2000;	2000US-0225270P.	
PR	14-AUG-2000;	2000US-0225447P.	
PR	14-AUG-2000;	2000US-0225757P.	
PR	14-AUG-2000;	2000US-0225758P.	
PR	14-AUG-2000;	2000US-0225759P.	
PR	18-AUG-2000;	2000US-0226279P.	
PR	22-AUG-2000;	2000US-0226681P.	
PR	22-AUG-2000;	2000US-0226868P.	
PR	22-AUG-2000;	2000US-0227182P.	
PR	23-AUG-2000;	2000US-0227009P.	
PR	30-AUG-2000;	2000US-0228924P.	
PR	01-SEP-2000;	2000US-0229287P.	
PR	01-SEP-2000;	2000US-0229343P.	
PR	01-SEP-2000;	2000US-0229344P.	
PR	01-SEP-2000;	2000US-0229345P.	
PR	05-SEP-2000;	2000US-0229509P.	
PR	05-SEP-2000;	2000US-0229513P.	
PR	06-SEP-2000;	2000US-0230437P.	
PR	06-SEP-2000;	2000US-0230438P.	
PR	08-SEP-2000;	2000US-0231242P.	
PR	08-SEP-2000;	2000US-0231243P.	
PR	08-SEP-2000;	2000US-0231244P.	
PR	08-SEP-2000;	2000US-0231413P.	
PR	08-SEP-2000;	2000US-0231414P.	
PR	08-SEP-2000;	2000US-0232080P.	
PR	08-SEP-2000;	2000US-0232081P.	
PR	12-SEP-2000;	2000US-0231968P.	
PR	14-SEP-2000;	2000US-0232397P.	
PR	14-SEP-2000;	2000US-0232398P.	
PR	14-SEP-2000;	2000US-0232399P.	
PR	14-SEP-2000;	2000US-0232400P.	
PR	14-SEP-2000;	2000US-0232401P.	
PR	14-SEP-2000;	2000US-0233063P.	
PR	14-SEP-2000;	2000US-0233064P.	
PR	21-SEP-2000;	2000US-0233065P.	
PR	21-SEP-2000;	2000US-0234223P.	
PR	21-SEP-2000;	2000US-0234274P.	
PR	25-SEP-2000;	2000US-0234997P.	
PR	25-SEP-2000;	2000US-0234998P.	
PR	26-SEP-2000;	2000US-0235484P.	
PR	27-SEP-2000;	2000US-0235834P.	
PR	27-SEP-2000;	2000US-0235836P.	
PR	29-SEP-2000;	2000US-0236327P.	
PR	29-SEP-2000;	2000US-0236367P.	
PR	29-SEP-2000;	2000US-0236368P.	
PR	29-SEP-2000;	2000US-0236369P.	
PR	29-SEP-2000;	2000US-0236370P.	
PR	02-OCT-2000;	2000US-0236802P.	
PR	02-OCT-2000;	2000US-0237037P.	
PR	02-OCT-2000;	2000US-0237038P.	
PR	02-OCT-2000;	2000US-0237039P.	
PR	02-OCT-2000;	2000US-0237040P.	
PR	13-OCT-2000;	2000US-0239935P.	
PR	13-OCT-2000;	2000US-0239937P.	
PR	20-OCT-2000;	2000US-0240960P.	
PR	20-OCT-2000;	2000US-0241221P.	
PR	20-OCT-2000;	2000US-0241785P.	
PR	20-OCT-2000;	2000US-0241786P.	
PR	20-OCT-2000;	2000US-0241787P.	
PR	20-OCT-2000;	2000US-0241808P.	
PR	20-OCT-2000;	2000US-0241809P.	
PR	20-OCT-2000;	2000US-0241826P.	
PR	01-NOV-2000;	2000US-0244617P.	
PR	08-NOV-2000;	2000US-0246474P.	

08-NOV-2000; 2000US-0246475P.
 08-NOV-2000; 2000US-0246476P.
 08-NOV-2000; 2000US-0246477P.
 08-NOV-2000; 2000US-0246478P.
 08-NOV-2000; 2000US-0246523P.
 08-NOV-2000; 2000US-0246524P.
 08-NOV-2000; 2000US-0246525P.
 08-NOV-2000; 2000US-0246526P.
 08-NOV-2000; 2000US-0246527P.
 08-NOV-2000; 2000US-0246528P.
 08-NOV-2000; 2000US-0246532P.
 08-NOV-2000; 2000US-0246609P.
 08-NOV-2000; 2000US-0246610P.
 08-NOV-2000; 2000US-0246611P.
 08-NOV-2000; 2000US-0246613P.
 17-NOV-2000; 2000US-0249207P.
 17-NOV-2000; 2000US-0249208P.
 17-NOV-2000; 2000US-0249209P.
 17-NOV-2000; 2000US-0249210P.
 17-NOV-2000; 2000US-0249211P.
 17-NOV-2000; 2000US-0249212P.
 17-NOV-2000; 2000US-0249213P.
 17-NOV-2000; 2000US-0249214P.
 17-NOV-2000; 2000US-0249215P.
 17-NOV-2000; 2000US-0249216P.
 17-NOV-2000; 2000US-0249217P.
 17-NOV-2000; 2000US-0249218P.
 17-NOV-2000; 2000US-0249244P.
 17-NOV-2000; 2000US-0249245P.
 17-NOV-2000; 2000US-0249264P.
 17-NOV-2000; 2000US-0249285P.
 17-NOV-2000; 2000US-0249297P.
 17-NOV-2000; 2000US-0249299P.
 17-NOV-2000; 2000US-0249300P.
 01-DEC-2000; 2000US-0250160P.
 01-DEC-2000; 2000US-0250391P.
 05-DEC-2000; 2000US-0251030P.
 05-DEC-2000; 2000US-0251988P.
 06-DEC-2000; 2000US-0256719P.
 08-DEC-2000; 2000US-0251479P.
 08-DEC-2000; 2000US-0251856P.
 08-DEC-2000; 2000US-0251868P.
 08-DEC-2000; 2000US-0251869P.
 08-DEC-2000; 2000US-0251989P.
 08-DEC-2000; 2000US-0251990P.
 11-DEC-2000; 2000US-0254097P.
 05-JAN-2001; 2001US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-476159/51.

P-PSDB; ADM19913.

Isolated nucleic acid molecule encoding a channel/transporter protein is used in preventing, treating or ameliorating a medical condition.

Claim 1; SEQ ID NO 241; 809pp; English.

The invention relates to an isolated nucleic acid molecule encoding a channel/transporter protein or sequences at least 95% identical to a channel. The nucleic acids and proteins encoded by them are used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathological condition. The antibodies to the proteins can also be used in alleviating symptoms associated with the disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays (ELISA). Disorders which are diagnosed or treated include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischemia, angiogenesis,

CC nervous system disorders e.g. Alzheimer's disease, infections caused by
 CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection.
 CC The polypeptides can also be used to aid wound healing and epithelial
 CC cell proliferation, to prevent skin aging due to sunburn, to maintain
 CC organs before transplantation, for supporting cell culture of primary
 CC tissues, to regenerate tissues and in chemotaxis. The polypeptides can
 CC also be used as a food additive or preservative to increase or decrease
 CC storage capabilities. This sequence corresponds to a gene of the
 CC invention.

SQ Sequence 1589 BP; 307 A; 412 C; 414 G; 456 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 2,61e-35 Length: 1589
 Score: 426.00 Matches: 84
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 34.1% Indels: 0
 DB: 5 Gaps: 0

US-10-073-293A-4 (1-245) x ADM19434 (1-1589)

Qy 1 MetGluSerProThrProGlnProAlaProGlySerAlaThrPheMetGluGlyCysLys 20
 Db 1338 ATGGAAAGCCCTACTCCACAGCCCTCTCTGGTTCGGCGACCTTCATGGAGGATGCATA 1397
 Qy 21 AspSerLeuProIleValIleSerTyrIleProValAlaPheAlaPheGlyLeuAsnAla 40
 Db 1398 GACAGTTTACCGATTGTTATTAGTTATATTCGGTGGCCCTTGGGTTCGGTCTGAATGCG 1457
 Qy 41 ThrArgLeuGlyPheSerProLeuGluSerValPhePheSerCysIleIleTyrAlaGly 60
 Db 1458 ACCCGTCTGGGATTCTCTCCTCTCGAAAGCGTCTTTTCTCTCTGCATCATTTATGCGAGC 1517
 Qy 61 AlaSerGlnPheValIleThrAlaMetLeuAlaGlySerSerIleuTyrIleAlaAla 80
 Db 1518 GCGAGCCAGTTCGTCATTACCGGATGCTGGCAGCGGAGTAGTTTGTGATTGCTGCA 1577
 Qy 81 LeuThrValMet 84
 Db 1578 CTGACCGTCATG 1589

RESULT 13

ABX58385

ID ABX58385 standard; cDNA; 1589 BP.

XX ABX58385;

XX 26-FEB-2003 (first entry)

XX cDNA encoding novel human musculoskeletal system antigen #729.

Gene; ss; musculoskeletal system antigen; cancer; metastasis;
 re-vascularisation; thrombosis; arteriosclerosis; mineral content;
 cardiovascular condition; wound; injury; burn; angiogenesis; ulcer;
 post-operative tissue repair; limb regeneration; neuronal growth;
 neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
 AIDS-related complex; chondrocyte growth; bone regeneration;
 periodontal regeneration; tissue transport; bone graft; skin aging;
 keratinocyte growth; hair loss; melanocyte growth; cell proliferation;
 cell growth; organ transplant; cell differentiation; body height; weight;
 hair colour; eye colour; skin; percentage of adipose tissue;
 pigmentation; cosmetic surgery; metabolism; biorhythm; cardiac rhythm;
 depression; tendency for violence; pain; reproductive capability;
 hormone level; endocrine level; appetite; libido; memory; stress;
 storage capability; fat content; lipid content; protein content;
 carbohydrate content; vitamin content; cofactor content;
 nutritional component.

XX Homo sapiens.

OS US2002147140-A1.

PN

XX

```
PD 10-OCT-2002.
XX
XX 17-JAN-2001; 2001US-00764877.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 28-JUN-2000; 2000US-0214886P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 22-AUG-2000; 2000US-0225758P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 21-SEP-2000; 2000US-0231413P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 27-SEP-2000; 2000US-0235834P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 17-NOV-2000; 2000US-0249299P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
XX
XX (ROSE/) ROSEN C A.
XX (RUBE/) RUBEN S M.
XX (BARA/) BARASH S C.
XX
XX Rosen CA, Ruben SM, Barash SC;
XX
XX WPI, 2003-128199/12.
XX P-FSDB; ABU13109.
XX
XX Isolated nucleic acid molecules encoding musculoskeletal system
XX associated polypeptides, useful for detecting disorders, e.g. cancer.
XX
XX Claim 1; SEQ ID NO 739; 321pp; English.
XX
XX The invention describes an isolated nucleic acid molecule comprising a
XX sequence encoding musculoskeletal system associated polypeptides useful
XX for detecting disorders, e.g., cancer or cancer metastases, in animals or
XX humans. The nucleic acid; stimulates re-vascularisation of ischaemic
XX
```

tissues associated with conditions such as thrombosis, arteriosclerosis, and other cardiovascular conditions; treats wounds due to injuries, burns, post-operative tissue repair, and ulcers; stimulates angiogenesis and limb regeneration; stimulates neuronal growth; can treat and prevent neuronal damage occurring in certain disorders or neurodegenerative conditions, such as, Alzheimer's disease, Parkinson's disease, and AIDS-related complex; stimulates chondrocyte growth, thus they can be used to enhance bone and periodontal regeneration and aid in tissue transports or bone grafts; prevents skin aging due to sunburn by stimulating keratinocyte growth; prevents hair loss, since FGF family members activate hair-forming cells and promotes melanocyte growth; stimulates growth and differentiation of hematopoietic cells and bone marrow cells when used in combination with other cytokines; maintains organs before transplantation or for supporting cell culture of primary tissues; induces tissue of mesodermal origin to differentiate in early embryos; increases or decreases the differentiation or proliferation of embryonic stem cells, besides, haematopoietic lineage; modulates mammalian characteristics, such as, body height, weight, hair colour, eye colour, skin, percentage of adipose tissue, pigmentation, size, and shape (e.g., cosmetic surgery); modulates mammalian metabolism; changes mammal's metal state or physical state by influencing biorhythms, circadian rhythms, depression, tendency for violence, tolerance for pain, reproductive capabilities, hormonal or endocrine levels, appetite, libido, memory, or stress; increases or decreases storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional components. This sequence encodes a novel human musculoskeletal system antigen. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the US patent office at
ftp.seqdata.uspto.gov/sequence.html?DocID=20020147140

XX SQ Sequence 1589 BP; 307 A; 412 C; 414 G; 456 T; 0 U; 0 Other;

Alignment Scores:
Pred. NO.: 2.61e-35 Length: 1589
Score: 426.00 Matches: 84
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 34.1% Indels: 0
DB: 8 Gaps: 0

US-10-073-293A-4 (1-245) x ABX58385 (1-1589)

Qy 1 MetGluSerProThrProGlnProAlaProGlySerAlaThrPheMetGluGlyCysIys 20
Db 1338 ATGGAAGCCCTACTCCACAGCCTGCTCTGGTTCGGGACCTTCATGGAAGGATGCAAA 1397
Qy 21 AspSerLeuProIleValIleSerTyrIleProValAlaPheAlaPheGlyLeuAsnAla 40
Db 1398 GACAGTTTACCGATTGTTATTAGTTATATCCGGTGGCCTTTGGTTCGGTCTGAATGCG 1457
Qy 41 ThrArgLeuGlyPheSerProLeuGluSerValPhePheSerCysIleIleIleTyrAlaGly 60
Db 1458 ACCGTCCTGGGATTCTCTCTCTCGAAAGCGTTTTTTTTTCTCCTCATCATTTATGAGGC 1517
Qy 61 AlaSerGlnPheValIleThrAlaMetIleAlaAlaGlySerSerLeuTrpIleAlaAla 80
Db 1518 GCGAGCCAGTTTCGTCAATTACCGCATGCTGCACGGGAGTAGTTTGTGGATTGCTGCA 1577
Qy 81 LeuThrValMet 84
Db 1578 CTGACCGTCATG 1589

RESULT 14
ADI53852
ID ADI53852 standard; cDNA; 1589 BP.
XX
AC ADI53852;
XX
XX 22-APR-2004 (first entry)
XX cDNA encoding novel human protein seq id 55.
XX

KW neuroprotective; nootropic; antiparkinsonian; anticonvulsant;
KW antidiabetic; antirheumatic; antiarthritic; dermatological;
KW antiinflammatory; immunosuppressive; antithyroid; antianemic; vasotropic;
KW anti-HIV; hepatotropic; virucide; antibacterial; fungicide;
KW antiparasitic; muscular; gynaecological; gastrointestinal; respiratory;
KW cardiovascular; antiarteriosclerotic; antiarrhythmic; cardiac;
KW nephrotropic; litholytic; cytostatic; gene therapy; neural disorder;
KW Alzheimer's disease; Parkinson's disease; Huntington's chorea;
KW amyotrophic lateral sclerosis; multiple sclerosis;
KW immune system disorder; diabetes; rheumatoid arthritis;
KW systemic lupus erythematosus; autoimmune thyroiditis; haemolytic anaemia;
KW inflammatory disorder; ischaemia-reperfusion injury;
KW inflammatory bowel disease; Crohn's disease; infectious disease;
KW HIV infection; hepatitis infection; bacterial infection;
KW fungal infection; parasitic infection; muscular disorder;
KW reproductive disorder; gastrointestinal disorder; pulmonary disorder;
KW cardiovascular disorder; atherosclerosis; arrhythmia; myocarditis;
KW renal disorder; acute glomerulonephritis; pyelonephritis;
KW renal lithiasis; proliferative disorder; cancerous diseases; human; gene;
KW ss.
OS Homo sapiens.
XX
XX
FN US2004018969-A1.
XX
PD 29-JAN-2004.
XX
PF 17-JAN-2001; 2001US-00764875.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0236366P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.

PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 11-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 XX
 FA (ROSE/) ROSEN C A.
 FA (RUBE/) RUBEN S M.
 FA (BARA/) BARASH S C.
 XX
 PI Rosen CA, Ruben SM, Barash SC;
 XX
 XX
 DR WPI; 2004-122079/12.
 DR P-PSDB; ADI54450.
 XX

New polypeptides and nucleic acid molecules, useful for detecting, preventing, diagnosing, prognosticating, treating or ameliorating medical conditions e.g. neural disorders, reproductive disorders or infectious diseases.

Claim 1; SEQ ID NO 55; 413pp; English.

The invention describes an isolated polypeptide comprising an amino acid sequence at least 90% identical to: a polypeptide fragment, domain, epitope, or full-length protein of any one of 607 amino acid sequences (I) described in the specification; a polypeptide fragment of (I), or the encoded sequence contained in (II), having biological activity; or a variant, allelic variant, or a species homologue of (I). The polypeptides and nucleic acid molecules are useful for detecting, preventing, diagnosing, prognosticating, treating or ameliorating medical conditions such as neural disorders, e.g. Alzheimer's disease, Parkinson's disease, Huntington's chorea, amyotrophic lateral sclerosis or multiple sclerosis,

Alignment Scores:
 Pred. No.: 2,61e-35 Length: 1589
 Score: 426.00 Matches: 84
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 34.1% Indels: 0
 DB: 12 Gaps: 0

US-10-073-293A-4 (1-245) x ADI53852 (1-1589)

QY 1 MetGluSerProThrProGlnProAlaProGlySerAlaThrPheMetGluGlyCysLys 20
 DB 1338 ATGGAAGCCCTACTCCAGCCTGCTCCTGTTCCGGACCTTCATCGAAGGATGCAAA 1397
 QY 21 AspSerLeuProIleValIleSerTyrIleProValAlaPheAlaPheGlyLeuAsnAla 40
 DB 1398 GACAGTTTACCGATTGTTATTAGTTATATCCGGTGGCCTTTGCGTCTCGATCGC 1457
 QY 41 ThrArgLeuGlyPheSerProLeuGluSerValPhePheSerCysIleIleTyrAlaGly 60
 DB 1458 ACCCGTCTGGGATCTCTCCCTCGAAGCGTTTTTTTCTCTGATCATTTATGCGAGC 1517
 QY 61 AlaSerGlnPheValIleThrAlaMetLeuAlaAlaGlySerSerLeuTrpIleAlaAla 80
 DB 1518 GCGAGCCAGTTCGTATTACCGCATGCTGGCAGCGGAGTAGTTTGTGGATTGCTGCA 1577

QY 81 LeuThrValMet 84
 DB 1578 CTGACCGTCATG 1589
 RESULT 15
 ID ADJ28112 standard; DNA; 1589 BP.
 AC ADJ28112;
 XX 20-MAY-2004 (first entry)
 XX Human musculoskeletal system-associated contig DNA - SEQ ID 739.
 DE musculoskeletal system; cytostatic; cancer; osteoporosis;
 KW gene therapy; vaccine; human; ds; gene.
 XX Homo sapiens.
 XX US2004009488-A1.
 XX 15-JAN-2004.
 XX 13-SEP-2002; 2002US-00242515.
 PR 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 24-FEB-2000; 2000US-0184664P.
 PR 02-MAR-2000; 2000US-0186350P.
 PR 16-MAR-2000; 2000US-0189874P.
 PR 17-MAR-2000; 2000US-0190076P.
 PR 18-APR-2000; 2000US-0198123P.
 PR 19-MAY-2000; 2000US-0205515P.
 PR 07-JUN-2000; 2000US-0209467P.
 PR 28-JUN-2000; 2000US-0214886P.
 PR 30-JUN-2000; 2000US-0215135P.
 PR 07-JUL-2000; 2000US-0216647P.
 PR 11-JUL-2000; 2000US-0217487P.
 PR 14-JUL-2000; 2000US-0217496P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 26-JUL-2000; 2000US-0220964P.
 PR 14-AUG-2000; 2000US-0224518P.
 PR 14-AUG-2000; 2000US-0224519P.
 PR 14-AUG-2000; 2000US-0225213P.
 PR 14-AUG-2000; 2000US-0225214P.
 PR 14-AUG-2000; 2000US-0225266P.
 PR 14-AUG-2000; 2000US-0225267P.
 PR 14-AUG-2000; 2000US-0225268P.
 PR 14-AUG-2000; 2000US-0225270P.
 PR 14-AUG-2000; 2000US-0225447P.
 PR 14-AUG-2000; 2000US-0225757P.
 PR 14-AUG-2000; 2000US-0225758P.
 PR 14-AUG-2000; 2000US-0225759P.
 PR 18-AUG-2000; 2000US-0226279P.
 PR 22-AUG-2000; 2000US-0226686P.
 PR 22-AUG-2000; 2000US-0227182P.
 PR 23-AUG-2000; 2000US-0227009P.
 PR 30-AUG-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 05-SEP-2000; 2000US-0229513P.
 PR 06-SEP-2000; 2000US-0230437P.
 PR 06-SEP-2000; 2000US-0230438P.
 PR 08-SEP-2000; 2000US-0231242P.
 PR 08-SEP-2000; 2000US-0231243P.
 PR 08-SEP-2000; 2000US-0231244P.
 PR 08-SEP-2000; 2000US-0231413P.

PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234224P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-024617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249255P.
PR 17-NOV-2000; 2000US-0249257P.

PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0251989P.
PR 06-DEC-2000; 2000US-025179P.
PR 06-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
PR 17-JAN-2001; 2001US-00764877.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM, Barash SC;
XX
XX WPI; 2004-090458/09.
XX P-PSDB; ADJ29135.
XX
XX New nucleic acid molecule, useful for preparing a medicament for preventing, treating or ameliorating a medical condition e.g., cancer of musculoskeletal tissues or osteoporosis.
XX
XX Claim 4; SEQ ID NO 739; 289pp; English.
XX
XX The invention relates to a novel isolated musculoskeletal system-associated nucleic acid molecule. The nucleic acid of the invention demonstrates cytostatic and osteopathic activities and may be useful for preparing a medicament for preventing, treating or ameliorating a medical condition such as cancer of the musculoskeletal tissues or osteoporosis, possibly via gene therapy or vaccine production. The current sequence is that of the human musculoskeletal system-associated contig DNA of the invention. The current sequence is not shown within the specification per se but is available on the USPTO web-site
XX http://seqdata.uspto.gov/sequence.html?DocID=20040009488.
XX
SQ Sequence 1589 BP; 307 A; 412 C; 414 G; 456 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2,61e-35 Length: 1589
Score: 426.00 Matches: 84
Percent Similarity: 100.0% Conservatives: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 34.1% Indels: 0
DB: 12 Gaps: 0

US-10-073-293A-4 (1-245) x ADJ28112 (1-1589)

Qy 1 MetGluSerProThrProGlnProAlaProGlySerAlaThrPheMetGluGlyCysLys 20
|||
Db 1338 ATGMAAGCCCTACTCCACAGCCTGCTCTGCTTCCGCGACCTTCATGGAAGGATGCAA 1397
|||
Qy 21 AspSerLeuProIleValIleSerTyrIleProValAlaIlePheGlyLeuAsnAla 40
|||
Db 1398 GACAGTTTACCGATTGTTATTAGTTATATTCGCGTCCGCTTGAATGCG 1457
|||
Qy 41 ThrArgLeuGlyPheSerProLeuGluSerValPhePheSerCysIleIleTyrAlaGly 60
|||
Db 1458 ACCCGTCGGGATTCCTCTCTCGAAGGGTTTTTTCTCTCGCATCATTTATGAGGC 1517
|||
Qy 61 AlaSerGlnPheValIleThrAlaMetLeuAlaGlySerSerLeuTyrIleAlaAla 80
|||
Db 1518 GCGAGCAGTTCGTGCTATTACCGCATGCTGCGAGCGGAGTAGTTTGTGATTGCTGCA 1577
|||
Qy 81 LeuThrValMet 84
|||
Db 1578 CTGACCGTCATG 1589

Search completed: February 16, 2006, 18:26:10
Job time : 769.065 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 16, 2006, 17:12:59 ; Search time 4989.47 Seconds
(without alignments)
2297.407 Million cell updates/sec

Title: US-10-073-293A-4
Perfect score: 1248
Sequence: 1 MESTPQAPGSAFMFGCK.....VCGCLTALIQAFWQGPDEL 245

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/abs/ABSSWEB.spool/US10073293/runat.15022006.120736.1008/app.query.fasta.1
-DB-EST -QFMT=fastap -SUPPFI=p2n.rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abs04
-USER=US10073293 @CGN 1.1 6731 @runat.15022006.120736.1008 -NCPU=6 -ICPU=3
-NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WAE TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*

1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_hic.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_est7.*
9: gb_gss1.*
10: gb_gss2.*
11: gb_gss3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	1248	100.0	770	10	CL689189 PRI0150C
C 2	564	45.2	396	11	Z88728 P.rubripes
C 3	224	17.9	516	9	A2130831 OSJNB007
C 4	127	10.2	326	9	A2578984 29g02 Sho
C 5	119	9.5	1648	2	BF240758 601875351
C 6	118	9.5	473	9	AQ936633 HSU41-905
C 7	111.5	8.9	807	9	A2211151 SP_0159_B

8	111.5	8.9	1118	10	AG396025
9	110.5	8.9	770	3	BJ574342
C 10	110.5	8.9	799	6	CF452672
C 11	109	8.7	864	10	AG127426
C 12	108.5	8.7	1067	3	BI767055
C 13	106.5	8.5	650	9	BH665038
C 14	106.5	8.5	812	9	BZ433690
C 15	106.5	8.5	1225	2	BI089645
C 16	106	8.5	1224	6	CB329965
C 17	105.5	8.5	744	6	CF445953
C 18	105	8.4	1162	8	CX117301
C 19	104.5	8.4	1163	2	BF981367
20	104	8.3	708	3	BJ576113
21	104	8.3	742	3	BJ578475
22	104	8.3	776	3	BJ566827
23	104	8.3	776	3	BJ573291
24	103.5	8.3	645	7	CK533251
25	103.5	8.3	1946	4	CNS0A5VA
C 26	103	8.3	880	9	AQ895122
C 27	103	8.3	973	7	CJ365181
C 28	103	8.3	1085	3	BM453898
C 29	103	8.3	1747	10	CL505761
C 30	102.5	8.2	775	3	BJ573793
C 31	102.5	8.2	1008	8	DN476416
32	102	8.2	689	9	CC147248
C 33	102	8.2	882	7	CV288706
C 34	102	8.2	922	7	CV289826
C 35	101.5	8.1	1797	10	AG279699
C 36	101	8.1	1009	4	AY109638
C 37	101	8.1	1304	8	DR121567
38	100.5	8.1	727	3	BJ573600
C 39	100.5	8.1	733	6	CF446315
C 40	100.5	8.1	746	5	BU128410
C 41	100.5	8.1	837	6	CD786623
C 42	100.5	8.1	1519	9	BZ579860
C 43	100	8.0	898	10	AG065695
C 44	100	8.0	918	10	AG072205
C 45	100	8.0	1232	2	BG426619

ALIGNMENTS

RESULT 1

CL689189/c

LOCUS

DEFINITION

pacificus var. California Pristionchus pacificus genomic, genomic

survey sequence.

CL689189

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBLISHED

COMMENT

CL689189 770 bp DNA linear GSS 09-JUL-2004
PRI0150C_C01 - PRI0150C BR (770) Mixed stage fosmid library of P.
pacificus var. California Pristionchus pacificus genomic, genomic
survey sequence.
CL689189 GI:50198486
GSS.
Pristionchus pacificus
Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Neodiplogasteridae; Pristionchus.
1 (bases 1 to 770)
Srinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.
AppADB: an AcedB database for the nematode satellite organism
Pristionchus pacificus
Nucleic Acids Res. 32 (1), D421-D422 (2004)
14681447
Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
This library was generated at Caltech, Pasadena, USA and end
sequenced at Vancouver, Canada.
Seq primer: T7
Class: fosmid ends.
Location/Qualifiers

```

source
1. .770
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Mixed stage fosmid library of P. pacificus
var. California"
/note="Vector: pEpifos-5 Fosmid vector"

ORIGIN
Alignment Scores:
Pred. No.: 2.24e-125 Length: 770
Score: 1248.00 Matches: 245
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 10 Gaps: 0

US-10-073-293A-4 (1-245) x CL689189 (1-770)

Qy 1 MetGluSerProThrProGlnProAlaProGlySerAlaThrPheMetGluGlyCysLys 20
Db 770 ATGGAAGCCCTACTCCACAGCTCTCTCGTTCGGCGACCTTCATGGAAGGATGCAAA 711
Qy 21 AspSerLeuProIleValIleSerTyrIleProValAlaPheAlaPheGlyLeuAsnAla 40
Db 710 GACAGTTTACCGATGTTATTAGTTATATTCCGGTGGCCCTTTCGGTCTCGAATGCG 651
Qy 41 ThrArgLeuGlyPheSerProLeuGluSerValPhePheSerCysIleIleTyrAlaGly 60
Db 650 ACCGCTCTGGGATCTCTCTCTCGAAGCGTTTTCCTCGATCATTTATGCGGC 591
Qy 61 AlaSerGlnPheValIleThrAlaMetLeuAlaAlaGlySerSerLeuTrpIleAlaAla 80
Db 590 GCGAGCCAGTTCGTTCATTACCGCATGCTGGCAGCCGGAGTAGTTTGTGATTCGTGCA 531
Qy 81 LeuThrValMetAlaMetAspValArgHisValLeuTyrGlyProSerLeuArgSerArg 100
Db 530 CTGACCGGCATGGCAATGGATGTTGCGCATGTTGTATGCGCGCTACTCGGTAGCCGT 471
Qy 101 IleIleGlnArgLeuGlnLysSerLysThrAlaLeuTrpAlaPheGlyLeuThrAspGlu 120
Db 470 ATTATTCAGCGCTTCGCAAAATCGAAACCCGCTGTGGGGTTCGCTGACGGATGAG 411
Qy 121 ValPheAlaAlaAlaThrAlaLysLeuValArgAsnAsnArgArgTrpSerGluAsnTrp 140
Db 410 GTTTTTCGCGCGCAACCGCAAACTGGTACGCAATAATCGCGCTGGAGCGAGAACTGG 351
Qy 141 MetIleGlyIleAlaPheSerSerTrpSerTrpValPheGlyThrValIleGlyAla 160
Db 350 ATGATCGGCATTCGCCCTTCAGTTTCATGTCGTGGGTATTTGGTACGGTAAATAGGGCA 291
Qy 161 PheSerGlySerGlyLeuLeuGlnGlyTyrProAlaValGluAlaAlaLeuGlyPheMet 180
Db 290 TTCTCCGCGAGCGGCTTCGTCGAGGTATCCGCGGTTGAAGCTGCATAGGTTTATG 231
Qy 181 LeuProAlaLeuPheMetSerPheLeuLeuAlaSerPheGlnArgLysGlnSerLeuCys 200
Db 230 CTTCGCGCACCTTTATGAGTTTCCTGCTCGCCTCTTTCGCGCGCAACAATCTCTTTCG 171
Qy 201 ValThrAlaAlaLeuValGlyAlaLeuAlaGlyValThrLeuPheSerIleProValAla 220
Db 170 GTTACCGCAGCGTATGTTGGTGCCCTTCGAGCGGTAACGCTATTTCCTATTCGCGTCGCC 111
Qy 221 IleLeuAlaGlyIleValCysGlyCysLeuThrAlaLeuIleGlnAlaPheTrpGlnGly 240
Db 110 ATTCTGGCAGCGATTTCTCTGGCTGCTCTACTGCGTTAATCCAGGCATTCTGGCAAGA 51
Qy 241 AlaProAspGluLeu 245
Db 50 GCGCCCGATGAGCTA 36

RESULT 2

```

```

FR0004918/c
LOCUS FR0004918 396 bp DNA linear GSS 25-FEB-2004
DEFINITION F.rubripes GSS sequence, clone 061111a8, genomic survey sequence.
ACCESSION Z88728
VERSION Z88728.1 GI:1865942
KEYWORDS GSS; genome survey sequence.
SOURCE Takifugu rubripes (Pugu rubripes)
ORGANISM Takifugu rubripes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontoidea; Tetraodontidae; Takifugu.
1
Elgar,G., Clark,M.S., Meek,S., Smith,S., Warner,S., Edwards,Y.J.,
Bouchireb,N., Cottage,A., Yeo,G.S., Umrانيا,Y., Williams,G. and
Brenner,S.
Generation and analysis of 25 Mb of genomic DNA from the pufferfish
Fugu rubripes by sequence scanning
Genome Res. 9 (10), 960-971 (1999)
10523524
2 (bases 1 to 396)
Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrانيا,Y.,
Williams,G. and Brenner,S.
Direct Submission
Submitted (18-FEB-1997) MRC Human Genome Mapping Project Resource
Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hgmrc.ac.uk
Vector: pBluescript II KS
V.type: phagemid
PRIMER: KS
DESCR:
One pass dye-terminator sequencing of cosmid cloned genomic
sequence.
Location/Qualifiers
1. .396
/organism="Takifugu rubripes"
/mol_type="genomic DNA"
/db_xref="taxon:31033"
/clone="061111a8"
/clone_lib="cosmid 061111"

ORIGIN
Alignment Scores:
Pred. No.: 3.68e-51 Length: 396
Score: 564.00 Matches: 113
Percent Similarity: 90.8% Conservative: 2
Best Local Similarity: 89.0% Mismatches: 12
Query Match: 45.2% Indels: 0
DB: 11 Gaps: 0

US-10-073-293A-4 (1-245) x FR0004918 (1-396)

Qy 1 MetGluSerProThrProGlnProAlaProGlySerAlaThrPheMetGluGlyCysLys 20
Db 396 GTGGAAGCCCTANTCCACAGCTGCTCCTGTCGAGACCTTCATGGAAGGATGCAAA 337
Qy 21 AspSerLeuProIleValIleSerTyrIleProValAlaPheAlaPheGlyLeuAsnAla 40
Db 336 GACAGTTTACCGATGTTATTAGTTATATTCCGGTGGCCCTTTCGGTCTGAATGCG 277
Qy 41 ThrArgLeuGlyPheSerProLeuGluSerValPhePheSerCysIleIleTyrAlaGly 60
Db 276 ACCGCTCTGGGATCTCTCTCTCGAAGCGTTTTCCTCGATCATTTATGCGGC 217
Qy 61 AlaSerGlnPheValIleThrAlaMetLeuAlaAlaGlySerSerLeuTrpIleAlaAla 80
Db 216 GCGAGCCAGTTCGTTCATTACCGCATGCTGGCAGCCGGAGTAGTTTGTGATTCGTGCA 157
Qy 81 LeuThrValMetAlaMetAspValArgHisValLeuTyrGlyProSerLeuArgSerArg 100
Db 156 CTGACCGGCATGGCAATGGATGTTGCGCATGTTGTATGCGCCCTACTCGGTAGCCGT 97
Qy 101 IleIleGlnArgLeuGlnLysSerLysThrAlaLeuTrpAlaPheGlyLeuThrAspGlu 120

```

Db 96 ATTATTCAGCTGCGAAAAATCGAAACCGCCCTGTGGGTGTACNGCCTGANGANGAG 37

Qy 121 ValPheAlaAlaAlaThrAla 127

Db 36 NTTTACTGCTGNAACCGCN 16

RESULT 3

AZ130831/c

LOCUS

DEFINITION

OSJNB0070K11r CUGI Rice BAC Library (EcoRI) Oryza sativa (japonica cultivar-group) genomic clone OSJNB0070K11r, genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

Wing,R.A. and Dean,R.A.

TITLE

A BAC End Sequencing Framework to Sequence the Rice Genome

JOURNAL

Unpublished (1998)

COMMENT

Contact: Wing RA

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: rwing@clemson.edu

Seq primer: GGAACAGCTATGACCATG

Class: BAC ends

High quality sequence start: 38

High quality sequence stop: 478.

FEATURES

source

1. 516

/organism="Oryza sativa (japonica cultivar-group)"

/mol_type="genomic DNA"

/cultivar="japonica"

/cultivar="Nipponbare"

/db_xref="taxon:3947"

/clone="OSJNB0070K11r"

/tissue type="Leaf"

/lab host="E. coli DH10B"

/clone lib="CUGI Rice BAC Library (EcoRI)"

/note="Vector: pBACIndigo; Site_1: EcoRI; Site_2: EcoRI; Rice is the most important food crop in the world. Half of the world population, especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety using EcoRI as the cloning enzyme. The library contains 55,296 clones with an average insert size of 121 kb providing approximately 15 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Three high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening and can be requested from the Clemson University BAC/EST Resource Center (www.genome.clemson.edu)."

ORIGIN

Alignment Scores:

Pred. No.: 6.63e-14 Length: 516

Score: 224.00 Matches: 43

Percent Similarity: 100.0% Conservative: 0

Best Local Similarity: 100.0% Mismatches: 0

Query Match: 17.9% Indels: 0

DB: 9 Gaps: 0

US-10-073-293A-4 (1-245) x AZ130831 (1-516)

Qy 1 MetGluSerProThrProGlnProAlaProGlySerAlaThrPheMetGluGlyCysLys 20

Db 131 ATGGAAAGCCCTACTCCACAGCCCTGCTCTGTTCCGCGACCTTCATCGAAGGATGCAAA 72

Qy 21 AspSerLeuProIleValIleSerTyrIleProValAlaPheAlaPheGlyIleuAsnAla 40

Db 71 GACAGTTTACCGATTGTTATTAGTTATATTCGCGTGGCTTTGCGTCTGAATCG 12

Qy 41 ThrArgLeu 43

Db 11 ACCCGTCTG 3

RESULT 4

AZ578984/c

LOCUS

DEFINITION

29g02 Shot-gun genomic library of Rhizobium strain ANU265 Rhizobium sp. NGR234 genomic clone 29g02, genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

Viprey,V., Rosenthal,A., Broughton,W.J. and Perret,X.

TITLE

Genetic snapshots of the Rhizobium species NGR234 genome

JOURNAL

Genome Biol. 1 (6), RESEARCH0014 (2000)

PUBMED

1178268

COMMENT

Contact: Virginie Viprey

Laboratoire de Biologie Moleculaire des Plantes Superieures

University of Geneva

1 Chemin de l'Imperatrice, Chambesy/Geneva 1292, Switzerland

Tel: +44 (0)1603450000

Fax: +44 (0)1603450045

Email: virginie.viprey@bbserc.ac.uk

Class: shotgun.

FEATURES

source

1. 326

/organism="Rhizobium sp. NGR234"

/mol_type="genomic DNA"

/strain="ANU265"

/db_xref="taxon:394"

/clone="29g02"

/clone lib="Shot-gun genomic library of Rhizobium strain ANU265"

/note="Vector: M13; derivative strain of NGR234 cured of pNGR234a"

ORIGIN

Alignment Scores:

Pred. No.: 0.00133 Length: 326

Score: 127.00 Matches: 37

Percent Similarity: 46.9% Conservative: 16

Best Local Similarity: 32.7% Mismatches: 53

Query Match: 10.2% Indels: 7

DB: 9 Gaps: 2

US-10-073-293A-4 (1-245) x AZ578984 (1-326)

Qy 51 ValPhePheSerCysIleIleTyrAlaGlyAlaSerGlnPheValIleThrAlaMetLeu 70

Db 326 GTGTTCATGAGCGCGCGGTCTATGCGGTGCAAGCAGATGTCGCGCATCGAGCTGTC 267

Qy 71 AlaAlaGlySerSerLeuTrpIleAlaAlaLeuThrValMetAlaMetAspValArgHis 90

Db 266 GGCAACAATGTCCAGCCCTGGCTGTGGTGTCTCGGTTTTCGCGGTCAACTTCGCGCAC 207

```

Qy 91 ValLeuTyrGlyProSerLeuArgSerArgile-----lleGlnArgLeuGlnLysSer 108
Db 206 GTGCTCTATTCCGGCTCGATCCCAACACATCCGTCATCTTCACGCTGGTCCAGAAG--- 150
Qy 109 LysThrAlaLeuTrpAlaPheGlyLeuThrAspGluValPheAlaAlaAlaThrAlaLys 128
Db 149 -----TTCTTCGCCTTTCTTCTGCTGATCGATCCGCAATATGCGGAGCGAAGG 99
Qy 129 LeuValArgAsnAsnArgArgTrpSerGluAsnTrpMetIleGlyIle-AlaPheSerSe 148
Db 98 CGGCGGAGCGGGCTTCGCTCAGCTTTCTTCGTTATCTCGGCTTCGGCTCGGTGTCATC 39
Qy 148 rTrpSerSerTrpValPheGlyThrValIleGlyAla 160
Db 38 TATTTCCTCCCTGCTCGTCAACACACGCGGACCGGCGCA 2

RESULT 5
BF240758/6 1648 bp mRNA linear EST 14-NOV-2000
LOCUS 601875351P1 NIH_MGC_55 Homo sapiens cDNA clone IMAGE:4091893 5',
DEFINITION mRNA sequence.
ACCESSION BF240758
VERSION BF240758.1 GI:11154682
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 1648)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLC948 row: e column: 14
High quality sequence stop: 484.
FEATURES
    source
        1..1648
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:4091893"
            /tissue_type="from acute myelogenous leukemia"
            /lab_host="DH10B (T1 phage-resistant)"
            /clone_lib="NIH MGC 55"
            /note="Organ: bone marrow; Vector: pDNR-LIB (Clontech);
            Site_1: SfII (ggcgccctcgccg); Site_2: SfiI
            (ggccattatggcc); Double-stranded cDNA was prepared from
            cell line RNA. 5' and 3' adaptors were used in cloning as
            follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and
            3' adaptor sequence:
            5'-ATTCTAGAGCGGCGCGCCGACATG-dT(30)BN-3' (where B = A,
            C, or G and N = A, C, G, or T). Average insert size
            1.65 kb (range 0.9-4.0 kb). 14/15 colonies contained
            inserts by PCR. This library was enriched for full-length
            clones and was constructed by Clontech Laboratories (Palo
            Alto, CA)."
```

ORIGIN

```

Alignment Scores:
Pred. No.: 0.0973 Length: 1648
Score: 119.00 Matches: 72
Percent Similarity: 41.1% Conservative: 36
Best Local Similarity: 27.4% Mismatches: 91
```

```

Query Match: 9.5% Indels: 64
DB: 2 Gaps: 13
US-10-073-293A-4 (1-245) x BF240758 (1-1648)
Qy 13 AlaThrPheMetGluGlyCysLysAspSerLeuProIleValIleSerTyrIleProVal 32
Db 1626 TCGGTTTTTCTTTCTGCGCTGC-----TTTCGTTGGTTTACTTAAATGTT 1585
Qy 33 AlaPheAlaPheGlyLeuAsnAlaThrArgLeuGlyPheSerProLeu----- 48
Db 1584 TCTGCTGATTTCCTCGCTCTTCTTCGACGTTTGGCAATCTTACCTATTCTCTACCGGCT 1525
Qy 49 -----GluSerValPhePheSer-CysIleIleTyrAlaGlyAlaSerGlnPhe-- 64
Db 1524 CCCTCTTCTCATGACTGCTTCTTATCGTTGCTCAITTACGCGTCTGTATATTCTC 1465
Qy 65 ----ValIleThr-AlaMetLeuAlaAlaGlySerSerLeuTrpIleAlaAla---LeuT 82
Db 1464 TCTGTTTCTCTTCTTCTCTGCTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTTCTTC 1405
Qy 82 hrValMetAlaMetAspValArgHisValLeuTyrGlyProSerLeuArgSerArgIleI 102
Db 1404 TCGTACTTTCGCTCAGTCTTCTTCTGCTGTTCTAT-----CCTTGGCGCTTCGTTCC 1351
Qy 102 leGlnArgLeuGlnLysSerIlyThrAlaLeuTrpAlaPheGlyLeuThrAspGluValP 122
Db 1350 TGTCTGCGCTCTCCCTATTGAGAGTGCCTCTGCTGCTCTCTCTCTCTCTCTCTCTCT 1300
Qy 122 heAlaAlaAlaThrAlaLysLeuValArgAsnAsnArgArgTrpSer---GluAsnTrpM 141
Db 1299 TCTCATCGCTCGCTCTCTCTGCTGCTGCTCTCTCTGCTGCTGCTCTCTCTCTCTCT 1243
Qy 141 etile-----GlyIleAlaPheSerSerTrpSerSerTrpValPheGlyT 156
Db 1242 GCCTCTTGGACCCCTCTCTCTGCTGCGCCCTCTCTCTATGTCCTGCTCTCTCTCTCT 1183
Qy 156 hrValIleGlyAlaPheSerGlySerGlyLeuLeuGlnGlyTyrProAlaValGluAla 176
Db 1182 TCGGCGTCTCTGCTCGGGTCTCGGCTCGCTCGCTCTCTCTCTCTCTCTCTCTCTCT 1129
Qy 176 laLeuGlyPheMetLeuProAlaLeuPheMetSerPhe-LeuLeuAla----- 191
Db 1128 TTCGGGCTCTC-----TCTCTCTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1078
Qy 192 -----SerPheGlnArgLysGlnSerLeuCysValThrAlaAlaLeuVal 206
Db 1077 CCGGTGCTCTGACCGTCTCTCTGCTGCTTTCGAGCTTTCGCGCTTCTGCTGCTGCTC 1018
Qy 207 Gly-AlaLeuAlaGlyValThrLeu----- 214
Db 1017 GGTGTGCTAGTCTGCTGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 958
Qy 215 -----PheSerIleProValAlaIleLeuAlaGlyIleValCysG1 228
Db 957 CGCATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 898
Qy 228 Y 228
Db 897 A 897

RESULT 6
AQ936633 473 bp DNA linear GSS 23-AUG-2000
LOCUS HSJ41-905 Human NotI clones Homo sapiens genomic, genomic survey
DEFINITION sequence.
ACCESSION AQ936633
VERSION AQ936633.1 GI:7213011
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
```

Hominidae; Homo.
 1 (bases 1 to 473)
REFERENCE
AUTHORS Zabarovsky, E.R., Gizatullin, R., Podowski, R.M., Zabarovska, V.V.,
 Xie, L., Muravenko, O.V., Kozyrev, S., Petrenko, L., Skobeleva, N.,
 Li, J., Protapopov, A., Kashuba, V., Ernberg, I., Winberg, G. and
 Wahlestedt, C.
TITLE NotI clones in the analysis of the human genome
JOURNAL Nucleic Acids Res. 28 (7), 1635-1639 (2000)
PUBMED 10710430
COMMENT Contact: Podowski RM
 Center for Genomics Research
 Karolinska Institute
 17177 Stockholm, Sweden
 Tel: +46-8-728-6372
 Fax: +46-8-337983
 Email: Raf.Podowski@cgr.ki.se
 Class: NotI site.
FEATURES
 source
 1..473
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone_lib="Human NotI clones"

ORIGIN
 Alignment Scores:
 Pred. No.: 0.0215 Length: 473
 Score: 118.00 Matches: 34
 Percent Similarity: 36.4% Conservative: 13
 Best Local Similarity: 26.4% Mismatches: 44
 Query Match: 9.5% Indels: 38
 DB: 9 Gaps: 1

US-10-073-293A-4 (1-245) x AQ936633 (1-473)

QY 8 ProAlaProGlySerAla----- 13
 Db 49 CCGGCACCGGGTCTCGCGGCGAGCAACACAGATCGTTCAAGCCAGCCACCTCAGC 108
 QY 13 ----- 13
 Db 109 CTCTGGCATGCTGCTACCGAACACAGTGTAGCCGCTCATGCCAGCGCGACACACC 168
 QY 14 -----ThrPheMetGluGlyCysLysAspSerLeuProIleValIleSerTyr 29
 Db 169 ATTGCCCGCCAGCGCTTCTTCACGGCGCCATCGCCATTTTGGCGTTATCCCTGGCGTC 228
 QY 30 IleProValAlaPheAlaPheGlyLeuAsnAlaThrArgLeuGlyPheSerProLeuGlu 49
 Db 229 GCCCTTGGGGCCTGCTGGCGGGTCCATCGCCATCGAGGCCAACCTTCAGCGCTGGGAG 288
 QY 50 SerValPhePheSerCysIleIleTyrAlaGlyAlaSerGlnPheValIleThrAlaMet 69
 Db 289 GGCAGGGGCTCTCGCCATCGTCNCGGGGCGCCGCGCATGTTGGTGGCCATCGGCATG 348
 QY 70 LeuAlaAlaGlySerSerLeuTrpIleAlaAlaLeuThrValMetAlaMetAspValArg 89
 Db 349 CTTAAAGCGGCGCAACCTGCTGTCGATCTGTGACCAACACATGCTGCTGACTTCGCAG 408
 QY 90 HisValLeuTyrGlyProSerLeuArg 98
 Db 409 CACCTGCTTACCGCTTGTCCATGCGC 435

RESULT 7
 AZ211151 807 bp DNA linear GSS 31-AUG-2000
LOCUS SP_0159_B1_F12_SP6E Strongylocentrotus purpuratus, purple sea
DEFINITION urchin, sperm genomic BAC library Strongylocentrotus purpuratus
 genomic clone Plate=159 Col=23 Row=L, genomic survey sequence.
 AZ211151
ACCESSION AZ211151.1 GI:8426644
VERSION GSS.
KEYWORDS Strongylocentrotus purpuratus
SOURCE

ORGANISM Strongylocentrotus purpuratus
 Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
 Echinoidea; Euechinoidea; Echinacea; Echinoida;
 Strongylocentrotidae; Strongylocentrotus.
REFERENCE
AUTHORS 1 (bases 1 to 807)
 Cameron, R.A., Mahairas, G., Rast, J.P., Martinez, P., Biondi, T.R.,
 Swartzell, S., Wallace, J.C., Poustka, A.J., Livingston, B.T.,
 Wray, G.A., Ettensohn, C.A., Lehrach, H., Britten, R.J., Davidson, E.H.
 and Hood, L.
TITLE A sea urchin genome project: Sequence scan, virtual map, and
 additional resources
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)
PUBMED 10920195
COMMENT Contact: Cameron, RA, Davidson, EH, Hood, L
 Division of Biology 156-29
 California Institute of Technology
 Pasadena California 91125, USA
 Tel: (626) 395-8421
 Fax: (626) 793-3047
 Email: acameron@caltech.edu
 Plate: 159 row: L column: 23
 Seq primer: SP6
 Class: BAC ends
 High quality sequence stop: 807.
FEATURES
 source
 1..807
 Location/Qualifiers
 /organism="Strongylocentrotus purpuratus"
 /mol_type="genomic DNA"
 /db_xref="taxon:7668"
 /clone_lib="Plate=159 Col=23 Row=L"
 /clone_lib="Strongylocentrotus purpuratus, purple sea
 urchin, sperm genomic BAC library"
 /notes="Organ: sperm; Vector: BACs3.6; BAC Clones in E-Coli
 DH10B"

ORIGIN
 Alignment Scores:
 Pred. No.: 0.234 Length: 807
 Score: 111.50 Matches: 66
 Percent Similarity: 35.0% Conservative: 24
 Best Local Similarity: 25.7% Mismatches: 85
 Query Match: 8.9% Indels: 82
 DB: 9 Gaps: 13

US-10-073-293A-4 (1-245) x AZ211151 (1-807)

QY 18 GlyCysLysAspSerLeu---ProIleValIleSerTyrIleProValAlaPheAlaPhe 36
 Db 126 GGCTGCGGTTCACTACTTTTTCATTTTGTATATACATTTTTCGTACGTTATAGAAAT 185
 QY 37 GlyLeuAsn-----AlaThrArgLeuGlyPheSerProLeuGluSer 50
 Db 186 TCTGCAATTCGTCATATTTTCTCTCTCGCCCTCTTTTCTCTCTCTCTCTCTCTCT 245
 QY 51 ValPhePheSerCysIleIleTyrAlaGlyAlaSerGlnPheValIleThr----- 67
 Db 246 TCCTTTTCTTCTGCTCTTATATATCTATCATCATCAGTTAATGTTGCTCTTATTATT 305
 QY 68 -----AlaMetLeuAlaAlaGlySerSer----- 75
 Db 306 ATCTTTAATGAGTAATCTTTTTCGCTTATCTGCTGTGGTGGTCTTCTCTCTCTCT 365
 QY 76 -----LeuTrpIleAlaAlaLeuThrValMetAlaMetAspValArgHis 90
 Db 366 GGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 395
 QY 91 ValLeuTyrGlyProSerLeuArgSerArgIleGlnArgLeuGlnLysSerLysThr 110
 Db 396 -----GGGGGGGCTGTGGGGTGTGGGGTGTGGGGTGTGGGGTGTGGGGTGTGG 446
 QY 111 AlaLeuTrpAlaPheGlyLeuThrAspGluValPheAlaAlaAlaAlaAlaAlaLysLeuVal 130
 Db 447 GTTGTGTGGCTCTTTTTCGTCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 497

```

Qy 131 ArgAsnAsnArgArgTrpSerGluAsnTrpMetIleGlyIleAlaPheSerSerTrpSer 150
Db 498 -----:-----:-----:-----:-----:-----:-----:
Qy 151 SerTrpValPheGlyThrValIleGlyAlaPheSerGlySerGlyLeuLeuGlnGlyTyr 170
Db 525 ---TGGGTGTTGGTTGGTCCGGG-----GGCTCGGG---CTGTCTGGGTTT 569
Qy 171 ProAlaValGluAlaAlaLeuGly-----PheMetLeuProAlaLeuPheMet 186
Db 570 GCGGCGCGCCCTGCCTCTGGCGTGGCGTGGCTGTCTCTGTTCCTCCGCTGTGTTTTC 629
Qy 187 SerPhe-LeuLeu-----:-----:-----:-----:AlaSerPheGlnAr 195
Db 630 TCTTCCCTCTCTGGCTCTCTTTCGCTGTGGTGACACTCCGACCTCGCTTCTTC 689
Qy 195 gLySGlnSerLeuCySValThrAlaAlaLeu---ValGlyAlaLeuAlaGlyValThrLe 214
Db 690 GCGGCGGTCTGGCCGTGGCGGTTCGCTGGGGCGGCGGTTCGGGGGTGTGTGTG 749
Qy 214 uPheSerIleProValAlaIleLeuAlaGlyIleValCysGlyCysLeu 230
Db 750 CCGCGCGCTGCC-----:-----:-----:-----:TTGGTCCGCGGTGTG 780

RESULT 8
AG396025 1118 bp DNA linear GSS 21-DEC-2004
LOCUS Mus musculus molossinus DNA, clone:MSMg01-215P11.T7, genomic survey
DEFINITION
ACCESSION AG396025
VERSION AG396025.1 GI:48028854
KEYWORDS GSS.
SOURCE Mus musculus molossinus (Japanese wild mouse)
ORGANISM Mus musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.

1
Abu,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
Shiroishi,T.
Contribution of Asian mouse subspecies Mus musculus molossinus to
genomic constitution of strain C57BL/6J, as defined by BAC-end
sequence-SNP analysis
Genome Res. 14 (12), 2439-2447 (2004)
15574823
2 (bases 1 to 1118)
Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
Direct Submission
Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa, 230-0045, Japan
(E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the mouse BAC library MSMg01. For BAC
library availability, please contact Kunya Abe (abe@crc.riken.jp).
Tsukuba Institute of Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@crc.riken.jp
PRIMERS
Sequencing : T7
LIBRARY : pBACe3.6
Vector : EcoRI
R.site 1 : EcoRI
R.site 2 : EcoRI.
location/Qualifiers
1..1118
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"

FEATURES
source

```

```

/db_xref="taxon:57486"
/clone="MSMg01-215P11.T7"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"

```

ORIGIN

```

Alignment Scores: 0.37 Length: 1118
Pred. NO.: 111.50 Matches: 70
Score: 111.50 Conservative: 36
Percent Similarity: 38.7% Mismatches: 98
Best Local Similarity: 25.5% Indels: 70
Query Match: 8.9% Gaps: 13
DB: 10

```

US-10-073-293A-4 (1-245) x AG396025 (1-1118)

```

Qy 4 ProThrProGlnProAlaProGlySerAla-----ThrPheMetGluGly 18
Db 55 CCTCTCCCGCGCGCCACCGCGCTTGTCTCCCGCGCGCGCTCGNTCGTTGGG 114
Qy 19 CysLysAspSerLeuProIleValIleSerTyrIleProValAla-----PheAla 35
Db 115 GGGGTGCGCTCTCTCCCGCTCTCTGCTGTGTGTGTGTGTGTGTGTGTGTGTG 174
Qy 36 PheGlyLeuAsnAlaThrArgLeuGlyPheSerProLeuGluSerValPhePheSer 54
Db 175 TTGGCCTCCCGGGG---CGTTGCGGTTTTTCGCTCTTTGTGTGTGTGTGTGTGTG 231
Qy 55 -----CysIleIleTyrAlaGlyAlaSerGlnPheValIleThrAlaMetLeuAla 71
Db 232 TCCGGTACGTGCTGGGTCTGTGGGTCTGTGGGTCTGTGTGTGTGTGTGTGTGTGT 291
Qy 72 AlaGlySerSerLeuTrpIleAlaAlaLeuThrValMetAlaMetAsp-ValArgHisVa 91
Db 292 TTGTTGTGTTCATGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGT 346
Qy 91 lLeuTyrGlyProSerLeuArgSerArgIleIleGlnArgLeuGlnLysSerLysThrAl 111
Db 347 -----CTCAGTGGCGTCTT-----:-----:-----:ACTCGTCCGG 372
Qy 111 aLeuTrpAlaPheGlyLeuThrAspGluValPheAlaAlaAlaThrAlaLysLeuValAr 131
Db 373 TGGTCTCGGTT-----:-----:-----:TTGCTGCCCTCTGGGCTGCTCTTTTGC 414
Qy 131 gAsnAsnArgArgTrpSerGluAsnTrpMetIleGlyIleAla-PheSerSerTrp 149
Db 415 TTTCGTTTGTGCGTGGTTGTGTGTGGGGGTTTGTGTGTGTGTGTGTGTGTGTGTGT 474
Qy 150 --SerSerTrpValPheGlyThrValIleGlyAlaPheSer-GlySerGlyLeuLeuGln 168
Db 475 TTTCGGGTGGGTGGTTCTTTGTCTCTGTTCCTGTTCCTGTTCCTGTTCCTGTTCCTGT 531
Qy 169 GlyTyrProAlaValGluAlaAlaLeuGlyPheMetLeuProAlaLeuPheMetSerPhe 188
Db 532 -----:-----:-----:-----:-----:-----:-----: 573
Qy 189 LeuLeuAlaSerPheGlnArgLysGlnSerLysCysValThrAlaAlaLeuValGlyAla 208
Db 574 -----TCTTTTCTCGC---TGTGGGCTTTGTGTGTGGGCGCTGGGTGTCCGCGG 621
Qy 209 LeuAlaGlyValThrLeuPheSerIleProValAlaIleLeuAlaGlyIle----- 225
Db 622 GCTGCTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 681
Qy 226 -----:-----:-----:-----:-----:-----:-----: 232
Db 682 CTTTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 741
Qy 233 Leu-----:-----:-----:-----:-----:-----:-----: 238
Db 742 CTGTGTGCTGTGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 777

```

RESULT 9

BJ574342 770 bp mRNA linear EST 18-DEC-2002
 LOCUS BJ574342 Ipomoea nil mixture of flower and flower bud Ipomoea nil
 DEFINITION cDNA clone jn26102 3', mRNA sequence.

ACCESSION BJ574342
 VERSION BJ574342.1 GI:27256170
 KEYWORDS EST.

SOURCE Ipomoea nil (Japanese morning glory)

ORGANISM

Ipomoea nil
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 asterids; lamiales; Solanales; Convolvulaceae; Ipomoeae; Ipomoea.

REFERENCE

1 (bases 1 to 770)
 Hoshino, A., Seki, M., Shin-i, T., Carninci, P., Kamiya, A., Shiraki, T.,
 Nitasaka, E., Shinozaki, K., Hayashizaki, Y., Kohara, Y. and Iida, S.
 ESTs of Japanese morning glory

AUTHORS

Unpublished (2002)

TITLE

Contact: Tadasu Shin-i

Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshini@genes.nig.ac.jp.

FEATURES

Source

1..770
 /organism="Ipomoea nil"
 /mol_type="mRNA"
 /cultivar="Tokyo-kokei standard"
 /db_xref="taxon:35883"
 /clone="jn26102"
 /tissue_type="mixture of flower and flower bud"
 /clone_lib="Ipomoea nil mixture of flower and flower bud"

ORIGIN

Alignment Scores:

Pred. No.: 0.282 Length: 770
 Score: 110.50 Matches: 52
 Percent Similarity: 36.1% Conservative: 23
 Best Local Similarity: 25.0% Mismatches: 74
 Query Match: 8.9% Indels: 59
 DB: 3 Gaps: 7

US-10-073-293A-4 (1-245) x BJ574342 (1-770)

QY 3 SerProThrProGlnProAlaProGlySerAlaThrPheMetGluGlyCysLysAspSer 22
 DB 166 ACCCTAAATCAAAACCGTATCCGCCACCGCTATCTATCTACCGCTTCCACATCACTTC 225
 QY 23 LeuProIleValIleSerTyrIleProValAlaPheAlaPheGlyLeuAlaThrArg 42
 DB 226 CTCCTCCCTTCCTCGCGCTGCTTCTTCTACCGGC---GACTTGACGGAC 270
 QY 43 LeuGlyPheSerProLeuGluSerValPhePheSerCysIleIleTyr----- 58
 DB 271 TTGGCTTCACGCCCTTACGGGAGCGCTTTTACAGGAGTAGCTTCTTCGGCGGACCC 330
 QY 59 -----AlaGlyAlaSerGlnPheValIleThrAlaMetLeuAla 71
 DB 331 TTGGCAGGAGGAGCGCTTTCCTCGGGTTGTATCGGGTGGCGCTCTCGGCAACCTTAGCC 390
 QY 72 AlaGlySerSerLeuTyrIleAlaAlaLeuThrValMetAlaMetAspValArgHisVal 91
 DB 391 GCCGGCTCCCTTCTTGGCTTGGCTGAGCGCTTAGTCTTCGACGTGCTCTTA----- 441
 QY 92 LeuTyrGlyProSerLeuArgSerArgIleIleIleGlnArgLeuGlnLysSerLysThrAla 111
 DB 442 -----GCCCTAGCCG 450
 QY 112 LeuTrpAlaPheGlyLeuThrAspGluValPheAlaAlaAlaThrAlaLysLeuValArg 131
 DB 451 GCAGCGCGCTTTGGCTTA-----GCCGCAACAGCC----- 480

QY 132 AsnAsnArgArgTTrpSerGluAsnTrpMetIleGlyIleAlaPheSerSerTrpSerSer 151
 DB 481 -----TTCCGCTTTGAGCTGGATTAGCC 504
 QY 152 TrpValPheGlyThrValIleGlyAlaPheSerGlySerGlyLeuLeuGlnGlyTyrPro 171
 DB 505 TTAGCAGCGGGCTTCGCTTCGGCTTGGCGCGGCTTTGGCTTGGCAGCAGCCTTG-- 561
 QY 172 AlavalGluAlaAlaLeuGlyPheMetLeuProAlaLeuPheMetSerPheLeuLeuAla 191
 DB 562 GCCTTAGCTTAGCTTGGCTTGGCAGGGGAGCAGCCTTCTTCTCTCTGGTAGCCTTT 621
 QY 192 SerPheGlnArgLysGlnSerLeu 199
 DB 622 GGCTTCGCCCGCGTCGACAGGCTT 645

RESULT 10

CF452672/c

LOCUS

DEFINITION

CF452672

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

1 (bases 1 to 799)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Expressed Sequence Tags from a normalized library of mixed onion

tissues (Allium cepa)

Unpublished (2003)

Contact: Hovey MJ

Department of Horticulture

USDA-ARS and University of Wisconsin

1575 Linden Drive, Madison, WI 53706, USA

Tel: 608-262-1830

Fax: 608-262-4743

Email: mjhavey@facstaff.wisc.edu

TIGR sequence name ACAEW85TR. For more information:

http://haveylab.hort.wisc.edu

Seq primer: CAG GAA ACA GCT ATG ACC.

Location/Qualifiers

1..799

/organism="Allium cepa"

/mol_type="mRNA"

/cultivar="Red Creole (bulbs), unknown (callus), Ebano &

Texas Legend (roots)"

/db_xref="taxon:4679"

/clone="ACAWE85"

/tissue_type="Callus, roots, and young bulbs"

/clone_lib="normalized cDNA library of onion"

/notes="Vector: pCMVSPORT6.1-ccdb (Invitrogen); Site 1:

ECORV (5'); Site 2: NotI (3'); Equal molar amounts of mRNA

from callus, roots, and young bulbs were combined to

synthesize the library. Normalization to enrich for

low-copy transcripts was performed by proprietary

techniques of Invitrogen."

ORIGIN

Alignment Scores:

Pred. No.: 0.297 Length: 799
 Score: 110.50 Matches: 63
 Percent Similarity: 43.6% Conservative: 29
 Best Local Similarity: 29.9% Mismatches: 69
 Query Match: 8.9% Indels: 51
 DB: 6 Gaps: 12

US-10-073-293A-4 (1-245) x CF452672 (1-799)

Db 304 TGCCTGGTGGGTTTTTTTTCGAAATCTTGTGCGCTTTCGGGGGGGTGGTGG-- 247

Qy 215 eSerileProValAlaLeuAlaLeuAlaGlyLeuValCysGlyCysLeuThrAlaLeuIleG1 235

Db 246 -----GTGGTTTGGGGGGGGTGGTTTGGTTGGTTGGTTGTTGCTTTTTTTT----- 202

Qy 235 nAlaPheTrp 238

Db 201 -----TTTTGG 196

RESULT 12

BI767055/c

LOCUS

DEFINITION

603054284F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5203652 5', mRNA sequence.

BI767055

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)

EST

BI767055.1 GI:15758633

1 (bases 1 to 1067)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

NIH-MGC <http://mgs.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Life Technologies, Inc.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Plate: LIAW1510 row: h column: 21

High quality sequence stop: 539.

Location/Qualifiers

1. .1067

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:5203652"

/lab_host="DH10B"

/clone_lib="NIH_MGC_122"

/note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source anonymous pool of 24 week female lung, 16 week female spleen, and 20-22 week male spleens. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 026. Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:

Pred. No.: 0.737 Length: 1067

Score: 108.50 Matches: 45

Percent Similarity: 34.7% Conservative: 23

Best Local Similarity: 23.0% Mismatches: 55

Query Match: 8.7% Indels: 73

DB: 3 Gaps: 7

US-10-073-293A-4 (1-245) x BI767055 (1-1067)

Qy 97 LeuArgSerArgIleIleGlnArgLeuGlnlySerLysThrAlaLeuTrpAlaPheGly 116

Db 934 TTACGACGAGCTATAGCAACGGCTTGAGAAAGTCGAAAGAGGTGCCGCCGTTAGAG 875

Qy 117 LeuThrAspGluValPheAlaAlaAlaThrAla----- 127

Db 874 GCCGGTCTTAAATTTTCGCTAGAGGTGGCGACACCCCAATAGGGTTGCCCC 815

Qy 128 -----LysLeuValAlaArgAsnAsnArgArgTrpSerGluAenTrp----- 140

Db 814 CCCCCAGAGGTAAAGGTGTCCCGTTCACACAGAGAGGTGGCGTGGGGTGGCCGAGG 755

Qy 141 -----MetIleGlyIleAlaPheSerSerTrpSerSerTrpValPheGlyThr 156

Db 754 TTGTTCCACCGTGCAGCGAGTGTGGGTTTAGCGTGGCGCGCGGTGGCGGTC 695

Qy 157 ValIleGlyAlaPheSerGlySerGlyLeuGlnGlyTyrProAlaValGluAla 176

Db 694 TTGTAT-----TTTCGCTCGCCCTCTTTGGGGGGGGCCACCCTTGGTACACAAAAA 641

Qy 177 LeuGly-----PheMetLeuPro----- 182

Db 640 CTGGGGGTAAACCTAAATTGTGGCCCCCTGGGGATATATTTTATGGCCCCCCCC 581

Qy 183 -----AlaLeuPheMetSerPheLeuLeuAlaSerPheGln 194

Db 580 CTTTTTTCTAGATTATATATATATATTTTGTGTTTATATAATCAATCTTTATTT 521

Qy 195 ArgLysGlnSerLeu----- 199

Db 520 AAAAAATCTAATCTGCCAGTTAGCGTTTTCACCACTCGGGGAGCTGAAACTTTTCA 461

Qy 200 -----CysValThrAlaAlaLeuValGlyAlaLeuAlaGlyValThrLeu 214

Db 460 GGCTTCACATCTTTTCTGCTGCTGCTTTGTAGTGCCTTAGCGCTTAGCA----- 413

Qy 215 PheSerIleProValAlaIleLeuAlaGlyIleValCysGlyCysLeu 230

Db 412 -----GCAGCCATTGTCAGTCTTTTATAGATGCTTGCTTA 380

RESULT 13

BH65038/c

LOCUS

DEFINITION

BOMHL31TR BO 2.3 KB Brassica oleracea genomic clone BOMHL31, genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Brassica oleracea

Brassica oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 650)

Ayele,M., Haas,B.J., Kumar,N., Wu,H., Xiao,Y., Van Aken,S., Utterback,T.R., Wortman,J.R., White,O.R. and Town,C.D.

Whole genome shotgun sequencing of Brassica oleracea and its application to gene discovery and annotation in Arabidopsis

Genome Res. 15 (4), 487-495 (2005)

15805490

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TR

Class: sheared ends.

Location/Qualifiers

1. .650

/organism="Brassica oleracea"

/mol_type="genomic DNA"

/strain="TO1000DH3"

/db_xref="taxon:3712"

/clone="BOMHL31"

/clone_lib="BO_2.3 KB"

/note="Vector: pROS1; Site_1: BstXI; 2-3 kb sheared

FEATURES

source

genomic DNA inserted into pHOS1 using BstXI linkers"

ORIGIN

Alignment Scores:
 Pred. No.: 0.606 Length: 650
 Score: 106.50 Matches: 45
 Percent Similarity: 44.2% Conservative: 23
 Best Local Similarity: 29.2% Mismatches: 54
 Query Match: 8.5% Indels: 32
 DB: 9 Gaps: 7

US-10-073-293A-4 (1-245) x BH665038 (1-650)

Qy 62 SerGlnPheValIleThrAlaMetLeuAlaAlaGlySerSerLeuTrpIleAlaAlaLeu 81
 Db 538 TCACGTGTTTATCTTACGAGCTGTAACATCAGCTGGTCTTCCGAGCTAATCCGGGTGTCA 479
 Qy 82 ThrValMetAlaMetAspValArgHisValLeuTyrGlyProSerLeuArgSerArgile 101
 Db 478 TCTCTTTCACCTACGACATCTACCGAACATACATAAATCTTAGAGCAACCGGAAA--- 422
 Qy 102 IleGlnArgLeuGlnLysSerLysThrAlaLeuTrpAlaPheGlyLeuThrAspGluVal 121
 Db 421 ---CAAATCTTGAAGTCTCAAGATCCGCGGTCTTGGATTTGGATGTTTCATCGGGATT 365
 Qy 122 PheAlaAlaThrAlaLysLeuValArgAsnAsnArgTrpSerGluAsnTrpMet 141
 Db 364 CTTGCA-----GTGATCTTGAACAAAGCCGGTGTTCCTCGGATGGATG 320
 Qy 142 -----ileGlyIleAlaPheSer---Ser 148
 Db 319 TATCTCGGATGGGAGTTCTCATTTGGCTCTCGGTTATTCCTCAATCGCTTTCATGCTTTTG 260
 Qy 149 TrpSer---SerTrpValPheGlyThrValIleGlyAlaPheSerGlySer-----Gly 165
 Db 259 TGGAGCAAGACTAATTCATTTGGTGTATTCTTGGTGCATCTCTGGCTGTGTTCTTTGGG 200
 Qy 166 Leu-LeuGlnGlyTyrPro-----AlaValGluAlaLeu 177
 Db 199 ATCGTTACTTGGTTATCCCGCAAGAAATTCAGTACGAGCGTGTGTGATCTTTGATACCACA 140
 Qy 177 uGlyPheMetLeuProAlaLeuPheMetSerPheLeuLeu 190
 Db 139 GGGAGAAATGCTCCCATCTGCTGGTAACTCTGTTGCTA 100

RESULT 14
 BZ433690/c
 LOCUS BZ433690 812 bp DNA linear GSS 13-DEC-2002
 DEFINITION genomic survey sequence.

ACCESSION BZ433690
 VERSION BZ433690.1 GI:26683096
 KEYWORDS GSS.
 SOURCE Brassica oleracea
 ORGANISM Brassica oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 rosids; eurosids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 812)
 Ayele.M., Haas,B.J., Kumar,N., Wu,H., Xiao,Y., Van Aken,S.,
 Utterback,T.R., Wortman,J.R., White O.R. and Town,C.D.
 Whole genome shotgun sequencing of Brassica oleracea and its
 application to gene discovery and annotation in Arabidopsis
 Genome Res. 15 (4), 487-495 (2005)

15805490
 Other GSSs: BONPR54TR
 Contact: Chris Town
 TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TF
 Class: sheared ends.
 FEATURES
 source Location/Qualifiers
 1..812

/organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /strain="TO1000DH3"
 /db_xref="taxon:3712"
 /clones="BONPR54"
 /clone_lib="BO.1.6.2 KB tot"
 /note="Vector: pHOS1; Site 1: BstXI; 1.6-2 kb sheared
 total DNA inserted into pHOS1 using BstXI linkers"

ORIGIN

Alignment Scores:
 Pred. No.: 0.829 Length: 812
 Score: 106.50 Matches: 45
 Percent Similarity: 44.2% Conservative: 23
 Best Local Similarity: 29.2% Mismatches: 54
 Query Match: 8.5% Indels: 32
 DB: 9 Gaps: 7

US-10-073-293A-4 (1-245) x BZ433690 (1-812)

Qy 62 SerGlnPheValIleThrAlaMetLeuAlaAlaGlySerSerLeuTrpIleAlaAlaLeu 81
 Db 473 TCACGTGTTTATCTTACGAGCTGTAACATCAGCTGGTCTTCCGAGCTAATCCGGGTGTCA 414
 Qy 82 ThrValMetAlaMetAspValArgHisValLeuTyrGlyProSerLeuArgSerArgile 101
 Db 413 TCTCTCTTCACTACGACATCTACCGAACATACATAAATCTTAGAGCAACCGGAAA--- 357
 Qy 102 IleGlnArgLeuGlnLysSerLysThrAlaLeuTrpAlaPheGlyLeuThrAspGluVal 121
 Db 356 ---CAAATCTTGAAGTCTCAAGATCCGCGGTCTTGGATTTGGATGTTTCATCGGGATT 300
 Qy 122 PheAlaAlaThrAlaLysLeuValArgAsnAsnArgTrpSerGluAsnTrpMet 141
 Db 299 CTTGCA-----GTGATCTTGAACAAAGCCGGTGTTCCTCGGATGGATG 255
 Qy 142 -----ileGlyIleAlaPheSer---Ser 148
 Db 254 TATCTCGGATGGGAGTTCTCATTTGGCTCTCGGTTATTCCTCAATCGCTTTCATGCTTTTG 195
 Qy 149 TrpSer---SerTrpValPheGlyThrValIleGlyAlaPheSerGlySer-----Gly 165
 Db 194 TGGAGCAAGCTAATTCATTTGGTGTATTCTTGGTGCATCTCTGGCTGTGTTCTTTGGG 135
 Qy 166 Leu-LeuGlnGlyTyrPro-----AlaValGluAlaLeu 177
 Db 134 ATCGTTACTTGGTTATCCCGCAAGAAATTCAGTACGAGCGTGTGTGATCTTGATACCACA 75
 Qy 177 uGlyPheMetLeuProAlaLeuPheMetSerPheLeuLeu 190
 Db 74 GGGAGAAATGCTCCCATCTGCTGGTAACTCTGTTGCTA 35

RESULT 15
 BI089645/c
 LOCUS BI089645 1225 bp mRNA linear EST 20-JUN-2001

DEFINITION 602854335F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:4995824 5',
 mRNA sequence.

ACCESSION BI089645
 VERSION BI089645.1 GI:14507975
 KEYWORDS EST.
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Hominidae; Homo.

1 (bases 1 to 1225)
 NIH-MGC http://mgs.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: c9apbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM11020 row: e column: 09
High quality sequence stop: 143.
Location/Qualifiers
1. 1225
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4995824"
/cell_line="MGC36"
/lab_host="DH10B"
/clone_lib="NTH_MGC 10"
/note="Organ: Cervix; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT
Average insert size 1.5 kb. Library prepared by Life Technologies."

This Page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus p2n model

Run on: February 16, 2006, 17:27:54 ; Search time 223.666 Seconds
(without alignments)
1947.114 Million cell updates/

Title: US-10-073-293A-4

Perfect score: 1248

Sequence: 1 MESPTQPAPGSATFMEGCK.....VCGCLTALIQAFWQGPDEL 245

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext
Ygapop 10.0 , Ygapext
Fgapop 6.0 , Fgapext
Delop 6.0 , Delext

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

```

-QC=frameset p2n.model -DEV=xlh
-Q/abes/ABSNWB spool/US10073293/runat 15022006 120738 1057/app_query.fasta_1
-DB=Issued Patents NA -QFMT=fascap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGNS=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MTNLEN=0 -MAXLEN=200000000
-HOST=abs08 -USER=US10073293 @CCN_1_1 237 @runat 15022006 120738 1057 -NCPU=6
-ICPU=3 -NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCAPOP=6 -FCAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

```

Database : Issued Patents NA:*

```

1: /cgm2_6/ptdata/1/ins/1 COMB.seq.*
2: /cgm2_6/ptdata/1/ins/5 COMB.seq.*
3: /cgm2_6/ptdata/1/ins/6A COMB.seq.*
4: /cgm2_6/ptdata/1/ins/6B COMB.seq.*
5: /cgm2_6/ptdata/1/ins/6 COMB.seq.*
6: /cgm2_6/ptdata/1/ins/PTCUR COMB.seq.*
7: /cgm2_6/ptdata/1/ins/PP COMB.seq.*
8: /cgm2_6/ptdata/1/ins/RE COMB.seq.*
9: /cgm2_6/ptdata/1/ins/backfiles1.seq.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1013.5	81.2	756	3	US-09-489-039A-3016	Sequence 3016, Ap
2	592	47.4	801	3	US-09-328-353-2399	Sequence 2399, Ap
3	246.5	19.8	744	3	US-09-489-039A-1307	Sequence 1307, Ap
4	229.5	18.4	593	3	US-09-710-273-267	Sequence 267, App
5	229.5	18.4	3058	3	US-09-710-273-3585	Sequence 3585, Ap
6	229.5	18.4	3077	3	US-09-710-273-3594	Sequence 3594, Ap
7	225.5	18.1	729	3	US-09-134-001C-2317	Sequence 2317, Ap
8	214.5	17.2	768	3	US-09-107-532A-1266	Sequence 1266, Ap
9	210.5	16.9	1038	3	US-09-252-991A-11376	Sequence 11376, A

ALIGNMENTS

RESULT 1

```

US-09-489-039A-3016
; Sequence 3016, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 3016
; LENGTH: 756
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-3016

```

Alignment Scores:	5.99e-116	Length:	756
Pred. NO.:	1013.50	Matches:	200
Score:	88.3%	Conservative:	18
Percent Similarity:	81.0%	Mismatches:	26
Best Local Similarity:	81.2%	Indels:	3
Query Match:	3	Gaps:	1
DB:			

US-10-073-293A-4 (1-245) x US-09-489-039A-3016 (1-756)

Qy	1	MetGluSerProThrProGlnProAla-----ProGlySerAlaThrPheMetGlu	17
		:::	
Db	10	ATGGAACCCCGGCCCTTAACTGTGCTTACCTGAACGGTGCACCGTCGGCGGA	69

Qy 18 GlyCysLysAspSerLeuProIleValIleSerThrIleProValAlaPheAlaPheCly 37
 Db 70 GCGCTCAAGACAGCTCTCCCATCGTCATCAGCTACCTTCGGTGGCATTCGCCCTTCGGA 129
 Qy 38 LeuAsnAlaThrArgLeuGlyPheSerProLeuGluSerValPhePheSerCysIleIle 57
 Db 130 CTCAATGCTACCCGCTGGGCTTACCCCGCTGGAAGCCCTGTCTTCTCTGCTGATATT 189
 Qy 58 TyrAlaGlyAlaSerGlnPheValIleThrAlaMetLeuAlaAlaGlySerSerLeuTrp 77
 Db 190 TAGCGCGCGCCAGCAGTTGTGATCACCGCCATGTTGGCGCGGCGAGTTGCTGTGG 249
 Qy 78 IleAlaAlaLeuThrValMetAlaMetAspValArgHisValLeuThrGlyProSerLeu 97
 Db 250 GTCGCGCGCCCTGACGCGTATGCGATGCGTCCGACAGCTGCTGATGGCCCTTCACATG 309
 Qy 98 ArgSerArgIleIleGlnArgLeuGlnLysSerLysThrAlaLeuTrpAlaPheGlyLeu 117
 Db 310 CGCAGCCGCATTCGTCAGCGCTGGATGATAAAGAAACCGCCCTTCGGCCCTG 369
 Qy 118 ThrAspGluValPheAlaAlaAlaThrAlaLysLeuValArgAsnAsnArgTrpSer 137
 Db 370 ACCGATGAGGTTTCGCGCGCTACCGCCAGGCTGTTCCGATAACCGCGCTGGAGC 429
 Qy 138 GluAsnTrpMetIleGlyIleAlaPheSerSerTrpSerSerTrpValPheGlyThrVal 157
 Db 430 GAAAGCTGATGCTCGGCTCGCTCTTTACCTCGTGGGCATCGTGGGTGCGGCACGCTG 489
 Qy 158 IleGlyAlaPheSerGlySerGlyLeuGlnGlyTyrProAlaValGluAlaAlaLeu 177
 Db 490 GCCGGCGCTGGTCCGGCAAGCGCTGCTGTGGATATCCGGCGGTGGAGCGCCCTC 549
 Qy 178 GlyPheMetLeuProAlaLeuPheMetSerPheLeuLeuAlaSerPheGlnArgLysGln 197
 Db 550 GGCCTTATGCTGCGCGCTGTTATGAGCTTCCTGCTGGCTCTCTTCCAGCGTCAGCAG 609
 Qy 198 SerLeuCysValThrAlaAlaLeuValGlyAlaLeuAlaGlyValThrLeuPheSerIle 217
 Db 610 TCACTCTGCTGACGGCGCGCTGCGCGCGCGCTGCGCGCGCTGCTGTTATTTTCCATT 669
 Qy 218 ProValAlaIleLeuAlaGlyIleValCysGlyCysLeuThrAlaLeuIleGlnAlaPhe 237
 Db 670 CCGCGCGCGATCTCCCGCGAATGTCTGCGGGTGTCTGACGGCGCTGCTTCAGGCCCATG 729
 Qy 238 TrpGlnGlyAlaProAspGlu 244
 Db 730 CTCAGGGGATGCGCGATGAA 750
 RESULT 2
 US-09-328-352-2399
 ; Sequence 2399, Application US/09328352
 ; Patent No. 6562958
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary L. Breton et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 ; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
 ; CURRENT APPLICATION NUMBER: US/09328,352
 ; PRIOR FILING DATE: 1999-06-04
 ; NUMBER OF SEQ ID NOS: 8252
 ; SEQ ID NO 2399
 ; LENGTH: 801
 ; TYPE: DNA
 ; ORGANISM: Acinetobacter baumannii
 US-09-328-352-2399
 Alignment Scores:
 Pred. No.: 1.29e-63 Length: 801
 Score: 592.00 Matches: 111
 Percent Similarity: 71.0% Conservative: 53
 Best Local Similarity: 48.1% Mismatches: 67
 Query Match: 47.4% Indels: 0

DB: 3 Gaps: 0
 US-10-073-293A-4 (1-245) x US-09-328-352-2399 (1-801)
 Qy 7 GlnProAlaProGlySerAlaThrPheMetGluGlyCysLysAspSerLeuProIleVal 26
 Db 76 CAGATGCTGCTCAACACAGCACTTTTGGCAAGTGCCAAAGATAGTCAAGCCATCGTT 135
 Qy 27 IleSerTyrIleProValAlaPheAlaPheGlyLeuAsnAlaThrArgLeuGlyPheSer 46
 Db 136 CTCACCTACTTACCACTTTCTTTTGGCTTTCTGTCATCTCAGTTTGGTTTACCA 195
 Qy 47 ProLeuGluSerValPhePheSerCysIleThrAlaGlyAlaSerGlnPheValIle 66
 Db 196 CCATGGGAGCATTTTCTTCTTGTTCATGATGAGTGCAGTGCAGCAATTTTGTAGTG 255
 Qy 67 ThrAlaMetLeuAlaAlaGlySerSerLeuTrpIleAlaLeuThrValMetAlaMet 86
 Db 256 GTCGCACTTTTAGCAAGTGGTCTTCTATATATGTTAAACCGCTTTAACAGTCATTGCACTT 315
 Qy 87 AspValArgHisValLeuThrGlyProSerLeuArgSerArgIleIleGlnArgLeuGln 106
 Db 316 GATATAGCATGTGACTTTTATGGGCTGCTCTTTATATATTTAAATTCACACTTAAATTAAT 375
 Qy 107 LysSerLysThrAlaLeuTrpAlaPheGlyLeuThrAspGluValPheAlaAlaAlaThr 126
 Db 376 TTAAGAAGACAGCCGTTTGGCTTGGGTTTAAAGACGAGTCTTGTGATGTTGGGCTTAAGTCTA 435
 Qy 127 AlalysLeuValArgAsnAsnArgArgTrpSerGluAsnTrpMetIleGlyIleAlaPhe 146
 Db 436 ATCCAACCTTTCACAGCGCAGACAACAATGGTCCGAGTCTTGGATGTTGGGCTTAAGTCTA 495
 Qy 147 SerSerTrpSerSerTrpValPheGlyThrValIleGlyAlaPheSerGlySerGlyLeu 166
 Db 496 TTTAGCTGGATGTCTGGCAAGTGGTTCATATTATTAGCGGGTATTTCAGATCAAGTCT 555
 Qy 167 LeuGlnGlyTyrProAlaValGluAlaLeuGlyPheMetLeuProAlaLeuPheMet 186
 Db 556 GCACACCTTCCAAAGTTTTTACAGCAGCTTTAGATTCTTATTACCTGCTTTATTTTAA 615
 Qy 187 SerPheLeuLeuAlaSerPheGlnArgLysGlnSerLeuCysValThrAlaAlaLeuVal 206
 Db 616 AGTTTTTTATAGCGCATTTGAACGAAAGCATAGCTTGTAGTTGCTGTATCGTTAGTG 675
 Qy 207 GlyAlaLeuAlaGlyValThrLeuPheSerIleProValAlaIleLeuAlaGlyIleVal 226
 Db 676 GTGTCAGCTTTGGCATGTATTGGAATTAACCTATCAGCGGCTATCTTTATTGGCATCTCT 735
 Qy 227 CysGlyCysLeuThrAlaLeuIleGlnAlaPhe 237
 Db 736 TCAGGCAATTTAGCAGGCTCTATTATTAATATTAT 768
 RESULT 3
 US-09-489-039A-1307
 ; Sequence 1307, Application US/09489039A
 ; Patent No. 6610836
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary Breton et. al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
 ; FILE REFERENCE: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
 ; CURRENT APPLICATION NUMBER: US/09489,039A
 ; PRIOR FILING DATE: 2000-01-27
 ; PRIOR APPLICATION NUMBER: US 60/117,747
 ; PRIOR FILING DATE: 1999-01-29
 ; NUMBER OF SEQ ID NOS: 14342
 ; SEQ ID NO 1307
 ; LENGTH: 744
 ; TYPE: DNA
 ; ORGANISM: Klebsiella pneumoniae
 US-09-489-039A-1307
 Alignment Scores:

```

Pred. No.: 8,4e-21 Length: 744
Score: 246.50 Matches: 67
Percent Similarity: 48.3% Conservative: 46
Best Local Similarity: 28.6% Mismatches: 106
Query Match: 19.8% Indels: 15
DB: 3 Gaps: 6

US-10-073-293A-4 (1-245) x US-09-489-039A-1307 (1-744)
Qy 18 GlyCysLysAspSerLeuProIleValIleSerTyrIleProValAlaPheAlaPheGly 37
Db 28 GCGCCAGCGTATATGCCCTATGATCGCATCGGACCTTCATTTTCATTTATTTGTCGCG 87
Qy 38 LeuAsnAlaThrArgLeuGlyPheSerProLeuGluSerValPhePheSerCysIleIle 57
Db 88 GCATTAAGCGTTAGCGCGGATGAGCGTCTGGCAGAGTACGGCGTGGTGGCGATCGTC 147
Qy 58 TyrAlaGlyAlaSerGlnPheValIleThrAlaMetIleuAlaAlaGlySerSerLeuTrp 77
Db 148 ATTGCCGCTCGCGCAAAATCTGGCACTGAATATGTTTAAACCGCGCCACGCTCGC 207
Qy 78 IleAlaAlaLeuThrValMetAlaMetAspValArgHisValLeuTyrGlyProSerLeu 97
Db 208 GTGATTATTTACACCATTAATATCAATCTTCCACGTCGTCGTATAGCGCTCTATT 267
Qy 98 ArgSerArgIleIleGlnArgLeuGlnIlyssSerLysThrAlaLeuTrpAlaPheGlyLeu 117
Db 268 ---TCCGCGACCGTGGCGAAGCTTCATTCTTTTAAATAATGCTTTATGTCCTATGCGCTC 324
Qy 118 ThrAspGluValPheAlaAlaAlaThrAlaLysLeuValArgAsnAsnArgArgTrpSer 137
Db 325 ACCGACGAGGTGTAT---GCCACACCGCTCAAGAGATGGGAAGAAATAAAGAGAAAA 381
Qy 138 GluAsnTrpMetIleGlyIleAlaPheSerSerTrpSerSerTrpValPheGlyThrVal 157
Db 382 TATCTGTTTATGTTGGCGATGATACCTCTGGCGGATTTGGGTCGTGGCTGATTTT 441
Qy 158 IleGlyAlaPheSerGlySerGlyLeuLeuGlnGlyTyrProAlaValGluAla---Ala 176
Db 442 CTTGGCGCGCTGGTGGCGGCC-----TCGTTTCCCAACATTGAGAAATACGGT 489
Qy 177 LeuGlyPheMetLeuProAlaLeuPheMetSerPheLeuLeuAlaSerPheGlnArgLys 196
Db 490 CTCGATTTGCGCATGGTGGCGCGCTTTATCGCCATCGTGGTGGCG-----CAAATATA 543
Qy 197 GlnSerLeuCysValThrAlaAlaLeuValGlyAlaLeuAlaGlyValThrLeuPheSer 216
Db 544 AGCAGCGCTGACCGTGGCGCGGTGTAGCCCGGTTTCCGCGCGTGTGTAGTGGTC 603
Qy 217 IlePro-----ValAlaIleLeuAlaGlyIleValCysGlyCysLeu 230
Db 604 CTACCTATTGCTGGGTATGCTGCTGCGCTCAGTGTGGCGTCTGGCGCTGGCGCTGTC 663
Qy 231 ThrAlaLeuIleGlnAlaPheTrpGlnGlyAlaProAspGlu 244
Db 664 GTCGACCTGGCGGAAGAGCGCAGCAGATGGCAAAACCGAA 705

RESULT 4
US-09-710-279-267
; Sequence 267, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; PRIOR FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 267
; LENGTH: 693

```

```

; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-09-710-279-267

Alignment Scores:
Pred. No.: 9,66e-19 Length: 693
Score: 229.50 Matches: 65
Percent Similarity: 49.3% Conservative: 47
Best Local Similarity: 28.6% Mismatches: 102
Query Match: 18.4% Indels: 13
DB: 3 Gaps: 6

US-10-073-293A-4 (1-245) x US-09-710-279-267 (1-693)
Qy 14 ThrPheMetGluGlyCysLysAspSerLeuProIleValIleSerTyrIleProValAla 33
Db 13 ACGTTTAAACAGAGGTGTAATCCACAGTATTCCACAGTATTCCAGGTGTAGGA 72
Qy 34 PheAlaPheGlyLeuAsnAlaThrArgLeuGlyPheSerProLeuGluSerValPhePhe 53
Db 73 CTATCGTTTGGAAATTTGGCGAGCTCTCCAAAATTTAGAGTATTTTATTATTG 132
Qy 54 SerCysIleIleTyrAlaGlyAlaSerGlnPheValIleThrAlaMetLeuAlaAlaGly 73
Db 133 TGCTGATTTATGCTGCTGCGAGCTCAATTTATTTTGTACATTAGTATGTCAGGC 192
Qy 74 SerSerLeuTrpIleAlaAlaLeuThrValMetAlaMetAspValArgHisValLeuTyr 93
Db 193 ACCCTATTTCTGCAATTTGCTTACAAATCTTATCGTAACTCTCGAATGTTCTTATTA 252
Qy 94 GlyProSerLeuArgSerArgIleIleGlnArgLeuGlnIlyssSerLysThrAlaLeuTrp 113
Db 253 AGTATGACTTTAGCACCTAATTAAGCAATATGGATTTTGGATAGGCTGGGCTTGA 312
Qy 114 AlaPheGlyLeuThrAspGluValPheAlaAlaAlaThrAlaLysLeuValArgAsnAsn 133
Db 313 ACGTTA---TTACAGATGAACCTTTTGGCGTGTCTATAACACCATATGTTAAA--GGT 366
Qy 134 ArgArgTrpSerGluAsnTrpMetIleGlyIleAlaPheSerSerTrpSerTrpVal 153
Db 367 GAAAAAATAAACGATCGATCGCTACACGAGTAAATATTAATCTACTTACTTATTTGGACT 426
Qy 154 PheGlyThrValIleGlyAlaPheSerGlySerGlyLeuGlnGlyTyrProAlaVal 173
Db 427 GTTTCCTGTGTAATCGGTGCCATTTTTCGGAGAG-----TATATTTCAAT 471
Qy 174 GluAlaAlaLeuGly-----PheMetLeuProAlaLeuPheMetSerPheLeuAla 191
Db 472 CCTGATGCGCTCGCGCTAGACTTTGCAATTTACCGCAATGTTATTTTATGTATATCT 531
Qy 192 SerPheGln-----ArgLysGlnSerLeuCysValThrAlaAlaLeuValGlyAlaLeu 209
Db 532 CAATTTGAAGGGAATAAGAAATCAGATTGAGAATATATATTGTTACTTCATTTGTTGTG 591
Qy 210 AlaGlyValThrLeuPhe-----SerIleProValAlaIleLeuAlaGlyIleValCys 227
Db 592 ATTGTGATGATGCTTCTTCAAGTCAATTTACCTTACCTATACCTAGCAATTTTAAATAGCC 651
Qy 228 GlyCysLeuThrAlaLeuIle 234
Db 652 GCAATGTTGCTGCATTTGTTA 672

RESULT 5
US-09-710-279-3585/c
; Sequence 3585, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US

```

; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3585
; LENGTH: 3058
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-09-710-279-3585

Alignment Scores:
Pred. No.: 1e-17 Length: 3058
Score: 229.50 Matches: 65
Percent Similarity: 49.3% Conservative: 47
Best Local Similarity: 28.6% Mismatches: 102
Query Match: 18.4% Indels: 13
DB: 3 Gaps: 6

US-10-073-293A-4 (1-245) x US-09-710-279-3585 (1-3058)

```
Qy 14 ThrPheMetGluGlyCysLysAspSerLeuProIleValIleSerTyrIleProValAla 33
Db 1397 ACGTTTAAACAAGGTGTGAAGAGTGTATCCCAAGTTACTTGGTTATGCGAGGTGTAGGA 1338
Qy 34 PheAlaPheGlyLeuAsnAlaThrArgLeuGlyPheSerProLeuGluSerValPhePhe 53
Db 1337 CTATCGTTTGAATGTGGCAGTCTCCCAAAATTCAGTGTTTAGAAATATTATTATTG 1278
Qy 54 SerCysIleIleTyrAlaGlyAlaSerGlnPheValIleThrAlaMetLeuAlaAlaGly 73
Db 1277 TGTCTGATTAATTTATGCTGTGGCAGCTCAATTTATTATTATTGTACATTAGTGTAGTCAGGC 1218
Qy 74 SerSerLeuTrpIleAlaAlaLeuThrValMetAlaMetAspValAcqHisValLeuTyr 93
Db 1217 ACCCTATTTCTGCAATTTGTGCTTACAATACTTATCGTTAACTCTCGAATGTTCTTATTA 1158
Qy 94 GlyProSerLeuArgSerArgIleGlnArgLeuGlnLysSerLysThrAlaLeuTrp 113
Db 1157 AGTATGACTTTAGCACCTAATTAATGAACAATATGATTTTGAATAGGTAGGCTTGA 1098
Qy 114 AlaPheGlyLeuThrAspGluValPheAlaAlaThrAlaLysLeuValArgAsnAsn 133
Db 1097 ACGTTA---TTAACAGATGAAACTTTTGGCGTTGCTATAACACCATATGTTTAAA---GGT 1044
Qy 134 ArgArgTrpSerGluAsnTrpMetIleGlyIleAlaPheSerSerTrpSerTrpVal 153
Db 1043 GAAAAAATTAACGATCGATGCGTACACGAGCTAAATATATCTGCTTACTTATTTGGACT 984
Qy 154 PheGlyThrValIleGlyAlaPheSerGlyLeuLeuGlnGlyTyrProAlaVal 173
Db 983 GTTTCCTGTGTAACGTCGTCATTTTCGAGAG-----TATATTCAAT 939
Qy 174 GluAlaAlaLeuGly-----PheMetLeuProAlaLeuPheMetSerPheLeuLeuAla 191
Db 938 CTGATGCGCTCGGCGTAGACTTTTGCAATTTACCGCAATGTTTATTTTATGATATATCT 879
Qy 192 SerPheGln-----ArgLysGlnSerLeuCysValThrAlaAlaLeuValGlyAlaLeu 209
Db 878 CAATTTGAAGGATTAAGAAATCAGATTGAGAAATATATATTTGTTACTCATTTGATGTGTG 819
Qy 210 AlaGlyValThrLeuPhe-----SerIleProValAlaIleLeuAlaGlyIleValCys 227
Db 818 ATTGTGATGATGCTTCTTCTAAGTTCAATTTCTACCTTCATACCTAGCAATTTTAATGCC 759
Qy 228 GlyCysLeuThrAlaLeuIle 234
Db 758 GCAATTTGTTGCTGCAATGTTA 738
```

RESULT 6

US-09-710-279-3594
; Sequence 3594, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3594
; LENGTH: 3077
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-09-710-279-3594

Alignment Scores:
Pred. No.: 1.01e-17 Length: 3077
Score: 229.50 Matches: 65
Percent Similarity: 49.3% Conservative: 47
Best Local Similarity: 28.6% Mismatches: 102
Query Match: 18.4% Indels: 13
DB: 3 Gaps: 6

US-10-073-293A-4 (1-245) x US-09-710-279-3594 (1-3077)

```
Qy 14 ThrPheMetGluGlyCysLysAspSerLeuProIleValIleSerTyrIleProValAla 33
Db 2396 ACGTTTAAACAAGGTGTGAAGAGTGTATCCCAAGTTACTTGGTTATGCGAGGTGTAGGA 2455
Qy 34 PheAlaPheGlyLeuAsnAlaThrArgLeuGlyPheSerProLeuGluSerValPhePhe 53
Db 2456 CTATCGTTTGAATTTGTGGCAGTCTCCCAAAATTTAGTGTTTAGAAATATTATTATTG 2515
Qy 54 SerCysIleIleTyrAlaGlyAlaSerGlnPheValIleThrAlaMetLeuAlaAlaGly 73
Db 2516 TGTCTGATTAATTTATGCTGTGGCAGCTCAATTTATTATTATTGTACATTAGTGTAGTCAGGC 2575
Qy 74 SerSerLeuTrpIleAlaAlaLeuThrValMetAlaMetAspValAcqHisValLeuTyr 93
Db 2576 ACCCTATTTCTGCAATTTGTGCTTACAATACTTATCGTTAACTCTCGAATGTTCTTATTA 2635
Qy 94 GlyProSerLeuArgSerArgIleGlnArgLeuGlnLysSerLysThrAlaLeuTrp 113
Db 2636 AGTATGACTTTAGCACCTAATTAATGAACAATATGATTTTGAATAGGTAGGCTTGA 2695
Qy 114 AlaPheGlyLeuThrAspGluValPheAlaAlaThrAlaLysLeuValArgAsnAsn 133
Db 2696 ACGTTA---TTAACAGATGAAACTTTTGGCGTTGCTATAACACCATATGTTTAAA---GGT 2749
Qy 134 ArgArgTrpSerGluAsnTrpMetIleGlyIleAlaPheSerSerTrpSerTrpVal 153
Db 2750 GAAAAAATTAACGATCGATGCGTACACGAGCTAAATATATCTGCTTACTTATTTGGACT 2809
Qy 154 PheGlyThrValIleGlyAlaPheSerGlyLeuLeuGlnGlyTyrProAlaVal 173
Db 2810 GTTTCCTGTGTAACGTCGTCATTTTCGAGAG-----TATATTCAAT 2854
Qy 174 GluAlaAlaLeuGly-----PheMetLeuProAlaLeuPheMetSerPheLeuLeuAla 191
Db 2855 CTGATGCGCTCGGCGTAGACTTTTGCAATTTACCGCAATGTTTATTTTATGATATATCT 2914
Qy 192 SerPheGln-----ArgLysGlnSerLeuCysValThrAlaAlaLeuValGlyAlaLeu 209
Db 2915 CAATTTGAAGGATTAAGAAATCAGATTGAGAAATATATATTTGTTACTCATTTGATGTGTG 2974
```


Qy 210 AlaGlyValThrLeuPhe-----SerIleProValAlaIleLeuAlaGlyIleValCys 227
Db 2975 ATTGTGATGATGCTCTCTTAAGTTCAATTCATCTTACCTTCATACCTAGCAATTTTAATAGCC 3034
Qy 228 GlyCysLeuThrAlaLeuIle 234
Db 3035 GCAATTGTTGTCATTTGTTA 3055

RESULT 7
US-09-134-001C-2317
; Sequence 2317, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2317
; LENGTH: 729
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2317

Alignment Scores:
Pred. No.: 3,28e-18 Length: 729
Score: 225.50 Matches: 66
Percent Similarity: 45.5% Conservative: 39
Best Local Similarity: 28.6% Mismatches: 89
Query Match: 18.1% Indels: 37
DB: 3 Gaps: 7

US-10-073-293a-4 (1-245) x US-09-134-001C-2317 (1-729)

Qy 14 ThrPheMetGluGlyCysLysAspSerLeuProIleValIleSerTyIleProValAla 33
Db 49 ACGTTTAAACAGAGTGTGAAGAGTATTCCTCCACGTTACTTGTGTGTTAGGTTAGGA 108
Qy 34 PheAlaPheGlyLeuAsnAlaThrArgLeuGlyPheSerProLeuGlySerValPhePhe 53
Db 109 CTATCGTTGGAAATGTGGAGCTCCCAAAATTCAGTGTTTAGAAATATTTTATTG 168
Qy 54 SerCysIleIleTyAlaGlyAlaSerGlnPheValIleThrAlaMetLeuAlaAlaGly 73
Db 169 TGTCTGATTTATTTATGCTGGTGCAGCCCAATTTATTTTGTACATTTAGTGTGTCAGGC 228
Qy 74 SerSerLeuTrpIleAlaLeuThrValMetAlaMetAspValArgHisValLeuTy 93
Db 229 ACCCTATTCTGCAATTTGTCTTACACACTTATCTGTTAACTCTCGAATGTTCTTATTA 288
Qy 94 GlyProSerLeuArgSerArgIleIleGlnArgLeuGlnLysSerLysThrAlaLeuTrp 113
Db 289 AGTATGACTTTAGCACCAATTAAGCAATATGATTTTGGAAATAGGGTAGGCTTGA 348
Qy 114 AlaPheGlyLeuThrAspGluValPheAlaAlaThrAlaLysLeuValArgAsnAsn 133
Db 349 ACGTTA---TTAACAGATGAACCTTTTGGAGTGTCTATAACGCCATATGTCAAA--GGT 402
Qy 134 ArgArgTrpSerGluAsnTrpMetIleGlyIleAlaPheSerSerTrpSerSerTrpVal 153
Db 403 GAAAAAATTAACGATCGATCGATGCTACACGCAATAATATTACTGCTTACTTATTTGGACT 462
Qy 154 PheGlyThrValIleGlyAlaPheSerGlySerGlyLeuLeuGlnGlyTyProAlaVal 173
Db 463 GTTTCATGTGTAATTCGGTGCATTTTCGGAGAG-----TATATTTCAAAT 507
Qy 174 GluAlaAlaLeuGly-----PheMetLeuProAlaLeuPheMetSerPheLeuLeuAla 191

Db 508 CCTGATCGCTGGCCCTAGACTTTGGCCATTACCGCAATGTTTATTTTATGATATATCT 567
Qy 192 SerPheGln-----ArgLysGlnSerLeu-----CysVal 201
Db 568 CAATTTCAGGGATTAGAAATACGATTCAGATATATATATTGTAATCTATTGATGTG 627
Qy 202 -----ThrAla 203
Db 628 ATTGTGATGATGCTTCTTCTAAGTTCAATTCATCTTACATACGATGCAATTTTAATAGCC 687
Qy 204 AlaLeuValGlyAlaLeuAlaGlyValThrLeu 214
Db 688 GCAATTGTTGCTGCAATTTAGGGTGTGTGATG 720

RESULT 8
US-09-107-532A-1266
; Sequence 1266, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; FILING DATE: 30-Jun-1998
; APPLICATION NUMBER: US/09/107,532A
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 1266:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 768 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...768
; SEQUENCE DESCRIPTION: SEQ ID NO: 1266:
US-09-107-532A-1266

Alignment Scores:
Pred. No.: 8,24e-17 Length: 768
Score: 214.50 Matches: 66
Percent Similarity: 46.8% Conservative: 43
Best Local Similarity: 28.3% Mismatches: 101

Query Match: 17.2% Indels: 23
DB: 3 Gaps: 6

US-10-073-293A-4 (1-245) x US-09-107-532A-1266 (1-768)

QY 20 LysAspSerLeuProIleValIleSerTyrlleProValAlaPheAlaPheGlyLeuAsn 39
DB 94 AAAGACAGTTACCTACCGTTTCGGTTATATCGGTATGACCTGCAATTTGGTATCGTT 153

QY 40 AlaThrArgLeuGlySerProLeuGlySerValPhePheSerCysIleIleTyrla 59
DB 154 GGGAAAGCTGCGGATTTTCATCCACTAGTCTGACGTTGATGTCCTGCTGCTATGCT 213

QY 60 GlyAlaSerGlnPheValIleThrAlaMetLeuAlaAlaGlySerSerLeuTrpIleAla 79
DB 214 GGTTCGCCCAATTTATCACAGTCAGCATGCTGCTAGTCACAGCCCAATTCCTTCCATC 273

QY 80 AlaLeuThrValMetAlaMetAspValArgHisValLeuTyrlGlyProSerLeu----- 97
DB 274 GTTTTCGACCTTCTAGTCAATTCGCCAATGATCTGATGAGTATGACGATGCTCCT 333

QY 98 -----ArgSerArgIleIleGlnArgLeuGlnLysSerLysThrAlaLeuTrpAla 114
DB 334 TATTTCAAGAAATCGGCTGCTTCAAAATCTG-----CTGATTGCG 375

QY 115 PheGlyLeuThrAspGluValPheAlaAlaAlaThrAlaLysLeuValArgAsnAsnArg 134
DB 376 ACATTGCTGCAGATGAAGTTTCGCTGCGAATGAATGAATGAATGAACTATCTGACAA 435

QY 135 ArgTrpSerGluAsnTrpMetIleGlyIleAlaPheSerSerTrpSerSerTrpValPhe 154
DB 436 AAGTTAACTTCGTTGGATGAATACACTAATGGATTTCTTATTTGACGTGGTGGGT 495

QY 155 GlyThrValIleGlyAlaPheSerGlySerGlyLeuGlnGlyTyrlProAlaValGlu 174
DB 496 TCTTCTCTGTTGGCGCACTTTAGGAAATTTTATTACTGAT-----CCGAAAAAG 546

QY 175 AlaAlaLeuGlyPheMetLeuProAlaLeuPheMetSer-----PheLeuLeuAlaSer 192
DB 547 TTTGGTTGGATTTGCGCATGCTGCTATGTTTATTCGGGTTTACTTTATCTTCAAGTCAAT 606

QY 193 PheGlnArgLysGlnSerLeuCysValThrAlaAlaLeuValGly-----AlaLeu 209
DB 607 TCTGATAGAACAACCTCTAAACGATTACAGTTGATCCTGATGTTGCTCCTCATTGATTCTG 666

QY 210 AlaGlyValThrLeuPheSerIlePro-----ValAlaIleLeuAlaGlyIle 225
DB 667 GTATATGTCGGACTGATTTTATCCCAAGCAACTTAGTGATTGTAGTTGTAACACTGATT 726

QY 226 ValCysGlyCysLeuThrAlaLeuIleGlnAlaPheTrp 238
DB 727 GGGTGGGATTTGGGGTGTGGATCAACATGCTTCTTT 765

RESULT 9

US-09-252-991A-11376
; Sequence 11376, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11376
; LENGTH: 1038
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-11376

Alignment Scores:
Pred. No.: 4.15e-16 Length: 1038
Score: 210.50 Matches: 73
Percent Similarity: 44.3% Conservative: 43
Best Local Similarity: 27.9% Mismatches: 108
Query Match: 16.9% Indels: 38
DB: 3 Gaps: 11

US-10-073-293A-4 (1-245) x US-09-252-991A-11376 (1-1038)

QY 1 MetGluSerProThrProGlnPro-AlaProGlySerAla-----Th 14
DB 297 CTGACCGCAGCTTCACCAACCGCCATGCCCCATGACATGCCACACCTGCCCCCAGC 356

QY 14 rPheMetGluGlyCysLysAspSerLeuProIleValIleSerTyrlleProValAlaPh 34
DB 357 CTTTCTCGGCGCGCGCTGCGCATCTCCCATTTGCTGCGCGCTGCGCGCTCGGGGCT 416

QY 34 eAlaPheGlyLeuAsnAlaThrArgLeuGlyPheSerProLeuGluSerValPhePheSe 54
DB 417 GCTGCGCGGATCGATGCGCCATCGAAGCCAACTCACCCTGCTGACCTCGCAGGCGCCAGGCGCTGTC 476

QY 54 rCysIleIleTyrlAlaGlyAlaSerGlnPheValIleThrAlaMetLeuAlaIaGlySe 74
DB 477 CGCATGCTCTTCGCCGCGCGCGCGCGCATGTTGGTGGCATGCGCATGCTCAAGGCGGTC 536

QY 74 rSerLeuTrpIleAlaAlaLeuThrValMetAlaMetAspValArgHisValLeuTyrlG 94
DB 537 CGGCTTCTTTTCGATCATCTTACACCTCTCTGACCTCTGACCTCGCAGCAGCTTCTACGG 596

QY 94 yProSerLeuArgSerArgIleIleGlnArgLeuGlnLysSerLysThrAlaLeuTrpAl 114
DB 597 CATGAGCTTGGCGCG-----GTCTGTCACTCTG-----CGGGAGCGTGGCG 641

QY 114 aPheGly-----LeuThrAspGluValPheAlaAlaAlaThrAlaLysLeuVa 130
DB 642 GATCGGCTGGGCTTCTGCTCACCGAGAGTTCTTGGCCCTCGCAGCAGCAGCGCG 701

QY 130 lArgAsnAsnArgArgTrpSerGluAsnTrpMetIleGlyIleAlaPheSerSerTrpSe 150
DB 702 GCGCAACTTCAACCGCTG-----TAGCCCTCGGCGTCTGCTGACCTTCTATAT 752

QY 150 rSerTrpValPheGlyThrValIleGlyAlaPheSerGlySerGlyLeuGlnGlyTy 170
DB 753 GCGCTGGAACCTGTTTACCTCGCGCGCATCTCTGCGCGCAGCATCTTGGCGCGCAGCAT----- 801

QY 170 rProAlaValGlu---AlaAlaLeuGlyPheMetLeuProAlaLeuPheMetSerPheLe 189
DB 802 -CCGGTCTCGAACAACCTGGGCGCTGGATTCTCCATCGCGCCACCTTCATCGCGCTGT 860

QY 189 uLeuAlaSerPheGlnArgLysGlnSerLeu---CysValThrAlaAlaLeuValGlyAl 208
DB 861 CGCGCGGCTGGTGGCAACGTAACGACACTGCTGCTGCTGCGCCACCTCGCTGTTCTGCTC 920

QY 208 aLeuAlaGlyValThrLeuPheSer-----IleProValAlaIle 221
DB 921 GGTG-----CTGTTTCAGCCACTGGCAATGGAGTTTCGGCGCTGCTGCGCGCG 968

QY 221 eLeuAlaGlyIleValCysGly-----CysLeuThrAlaLeuIleGlnAlaPheTrpGl 239
DB 969 TCTCGCGGATGGCGCGGTTTCACTGTCACAAAGCTGTATCGGAGGCGCCATGCTC 1028

QY 239 nGly 240
DB 1029 TGGG 1032

RESULT 10

US-09-252-991A-11418
; Sequence 11418, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:

[illegible]

```
Qy 34 PheAlaPheGlyLeuAenAlaThrArgLeuGlyPheSerProLeuGluSerValPhePhe 53
Db 1166 ATTCAATTGGTATTGGCTTCGCTCAAAACCTTAGTATTTAGAAATTTGCTTGTTA 1107
Qy 54 SerCysIleIleTyrAlaGlyAlaSerGlnPheValIleThrAlaMetLeuAlaAlaGly 73
Db 1106 TGTCTTGTTATATATGCGGCTGCTGCGCAATTTATTATGTCGGCTTTTATAGCAGGT 1047
Qy 74 SerSerLeuTrpIleAlaAlaLeuThrValMetAlaMetAspValArgHisValLeuTyr 93
Db 1046 ACACCTATATCAGCGATTGTACTACTGTATTTATCGTAAATTCAGAAATGCTCTTTTA 987
Qy 94 GlyProSerLeuArgSerArgIleIleGlnArgLeuGlnLysSerLysThr--AlaLeu 112
Db 986 AGTAGTCGCTTGCACCAAACTTC-----AAGACATATGGGGTTT 948
Qy 113 Trp---AlaPheGlyLeu-----ThrAspGluValPheAlaAlaAlaThrAla 127
Db 947 TGGAAACCGTGTGGATTAGGTTTCATTAGTAACAGCAACGTTTGGCGTCGCCATTACA 888
Qy 128 LysLeuValArgAsnAsnArgArgTrpSerGluAsnTrpMetIleGlyIleAlaPheSer 147
Db 887 CCTTATTAAAA---GGAGAAAGCTATCAATGATCGTTCGATCGATGGTCTTAAACATCACA 831
Qy 148 SerTrpSerSerTrpValPheGlyThrValIleGlyAlaPheSerGlySerGlyLeuLeu 167
Db 830 GCATATTATTTTGGGCAATTTTCATGTGTAGCTGGGCTTTATTGGCGAA-----780
Qy 168 GlnGlyTyrProAlaValGluAlaLeuGly-----PheMetLeuProAlaLeuPhe 185
Db 779 -----TATATCTCAATCCGAAACGCTAGAGTTTGTATTCGTAAGTATGTTTATGCGCTTCA 609
Qy 204 AlaLeuValGlyAlaLeuAlaGlyValThrLeuPheSerIleProVal-----Ala 220
Db 665 GTACTCATTTATGGCGTCATA---GTAATGATGTTATCGTAAGTATGTTTATGCGCTTCA 609
Qy 221 IleLeuAlaGlyIleValCysGlyCysLeuThrAlaLeuIle 234
Db 608 TATCTAGCAATATTAATTGCAGCCACAATTTTCAGCAGCGTTA 567
```

RESULT 12

```
US-08-781-986A-136/c
; Sequence 136, Application US/08781986A
; Patent No. 6737248
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:
; ADDRESS: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,986A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
```

```
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 136:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11823 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-781-986A-136

Alignment Scores:
Pred. No.: 1.06e-13 Length: 11823
Score: 204.50 Matches: 65
Percent Similarity: 47.0% Conservative: 45
Best Local Similarity: 27.8% Mismatches: 97
Query Match: 16.4% Indels: 27
DB: 3 Gaps: 10
```

US-10-073-293A-4 (1-245) x US-08-781-986A-136 (1-11823)

```
Qy 14 ThrPheMetGluGlyCysLysAspSerLeuProIleValIleSerTyrIleProValAla 33
Db 1226 AGTTTTAGACAGCGCTGCAAGAGTGTATCCACATATTGGGTATGCCGGTGGT 1167
Qy 34 PheAlaPheGlyLeuAenAlaThrArgLeuGlyPheSerProLeuGluSerValPhePhe 53
Db 1166 ATTTCAATTGGTATTGGCTTCGCTCAAACTTAGTATTTAGAAATTTGCTTGTTA 1107
Qy 54 SerCysIleIleTyrAlaGlyAlaSerGlnPheValIleThrAlaMetLeuAlaAlaGly 73
Db 1106 TGTCTTGTTATATATGCGGCTGCGCAATTTATTATGTCGGCTTTTATAGCAGGT 1047
Qy 74 SerSerLeuTrpIleAlaAlaLeuThrValMetAlaMetAspValArgHisValLeuTyr 93
Db 1046 ACACCTATATCAGCGATTGTACTACTGTATTTATCGTAAATTCAGAAATGCTCTTTTA 987
Qy 94 GlyProSerLeuArgSerArgIleIleGlnArgLeuGlnLysSerLysThr--AlaLeu 112
Db 986 AGTAGTCGCTTGCACCAAACTTC-----AAGACATATGGGGTTT 948
Qy 113 Trp---AlaPheGlyLeu-----ThrAspGluValPheAlaAlaAlaThrAla 127
Db 947 TGGAAACCGTGTGGATTAGGTTTCATTAGTAACAGCAACGTTTGGCGTCGCCATTACA 888
Qy 128 LysLeuValArgAsnAsnArgArgTrpSerGluAsnTrpMetIleGlyIleAlaPheSer 147
Db 887 CCTTATTAAAA---GGAGAAAGCTATCAATGATCGTTCGATCGATGGTCTTAAACATCACA 831
Qy 148 SerTrpSerSerTrpValPheGlyThrValIleGlyAlaPheSerGlySerGlyLeuLeu 167
Db 830 GCATATTATTTTGGGCAATTTTCATGTGTAGCTGGGCTTTATTGGCGAA-----780
Qy 168 GlnGlyTyrProAlaValGluAlaLeuGly-----PheMetLeuProAlaLeuPhe 185
Db 779 -----TATATCTCAAAATCCGAAACGCTAGAGTTTGTATTCACGCGCTATGTTT 726
Qy 186 MetSerPheLeuLeuAlaSerPheGln-----ArgLysGlnSerLeuCysValThrAla 203
Db 725 ATCTTTTGGCCATGCGCAATTTGAATCAATTAATCGCATTAATCGCATTAAGAAATTACATA 666
Qy 204 AlaLeuValGlyAlaLeuAlaGlyValThrLeuPheSerIleProVal-----Ala 220
Db 665 GTACTCATTTATGGCGTCATA---GTAATGATGTTATCGTAAGTATGTTTATGCGCTTCA 609
Qy 221 IleLeuAlaGlyIleValCysGlyCysLeuThrAlaLeuIle 234
Db 608 TATCTAGCAATATTAATTGCAGCCACAATTTTCAGCAGCGTTA 567

RESULT 13
```

```

US-09-328-352-360
; Sequence 360, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; PRIOR FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 360
; LENGTH: 792
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-360

Alignment Scores:
Pred. No.: 6,28e-15 Length: 792
Score: 199.50 Matches: 65
Percent Similarity: 47.3% Conservatives: 41
Best Local Similarity: 29.0% Mismatches: 101
Query Match: 16.0% Indels: 17
DB: 3 Gaps: 6

US-10-073-293A-4 (1-245) x US-09-328-352-360 (1-792)

Qy 15 PheMetGluGlyCysLysAspSerLeuProIleValIleSerTyIleProValAlaPhe 34
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 133 TTTTACGAGGGGGCGATAGATATATGCGCATCTATCTATTTCCGTTATACCATGGCGTATT 192

Qy 35 AlaPheGlyLeuAsnAlaThrArgLeuGlyPheSerProLeuGluSerValPhePheSer 54
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 193 TTAGCTGGATCAATGGCAGTTTCATGCGGTTTATCATCTTTTATAAGCAGCATGCGATGCA 252

Qy 55 CysIleIleTyAlaGlyAlaSerGlnPheValIleThrAlaMetLeuAlaAalaGlySer 74
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 253 GGAATTCTGTTGTGCGCGCTGCGCAATTAGTAAGTTTAAAGCATGGTGATGGAAGGGGCA 312

Qy 75 SerLeuTrpIleAlaAlaLeuThrValMetAlaMetAspValArgHisValLeuTyGly 94
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 313 TCGTTATTAACCATCTATGTCACATCTTTCTTCTTCTGCTCAACATTTTATTATGCT 372

Qy 95 ProSerLeuArgSerArgIleIleGlnArgLeuGlnLysSerIleThrAlaLeuTrpAla 114
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 373 TTAACCTTAAAGAAATGAC---ATTCTATTGCTCTCTCATAAAGATTAACCTTAGGT 429

Qy 115 PheGlyLeuThrAspGluValPheAlaAlaAlaThrAlaLysLeuValArgAsnArg 134
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 430 TTTCTGCTTACCGCAGAAATTATTTGCAGTAAGTGTCTCT-----AATGAG 474

Qy 135 ArgTrpSerGluAsnTrpMetIleGlyIleAlaPheSerSerTrpSerSerTrpValPhe 154
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 475 AAAAACAATCTCAATATCTATTCGAGCAGGTCTTTGCTTTTATTTATTTTGGGTGTT 534

Qy 155 GlyThrValIleGlyAlaPheSerGlySerGlyLeu-----LeuGlnGlyTyProAla 172
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 535 TTTAGTTTGTGTTATTTTATTTAGCAACGCGCTACCTAATCTATTAAATTTAT----- 588

Qy 173 ValGluAlaAlaLeuGlyPheMetLeuProAlaLeuPheMetSerPheLeuLeuAlaSer 192
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 589 -----CATTTGGATTTTTCATTTATGCTATTTTGTGTCATGATTTGTTCCCATG 639

Qy 193 PheGlnArgLysGlnSerLeuCysValThrAlaAlaLeuValGlyAlaLeuAlaGlyVal 212
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 640 TGTAAAGGGAGCGCT-----GTAAATGGCAGGCAATTTTAAATGCTGTGTCAGTGGTTT 693

Qy 213 ThrLeu-----PheSerIleProValAlaIleLeuAlaGlyIleValCysGlyCysLeu 230
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 694 GTTCTTAAATTTTTCATATAGAGGTGCCATCTTAATTTCTGTTTATTTGGGGATGTTT 753

Qy 231 ThrAlaLeuIle 234
|||:|||||:

```

```

Db 754 ATTCGCGTAATA 765
RESULT 14
US-09-471-803A-2
; Sequence 2, Application US/09471803A
; Patent No. 6613545
; GENERAL INFORMATION:
; APPLICANT: KENNERKNECHT, NICOLE
; APPLICANT: SAHM, HERMANN
; APPLICANT: EGGELE, LOTHAR
; APPLICANT: PEPPERLE, WALTER
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE CODING FOR THE EXPORT OF
; TITLE OF INVENTION: BRANCHED CHAIN AMINO ACIDS, PROCESS FOR THE
; TITLE OF INVENTION: ISOLATION THEREOF AND USE THEREOF
; FILE REFERENCE: 21123/265496/MAS
; CURRENT APPLICATION NUMBER: US/09/471,803A
; CURRENT FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: DE 199 51 708.8
; PRIOR FILING DATE: 1999-10-27
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 753
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(753)
; OTHER INFORMATION: drmp
; OTHER INFORMATION: ATCC14752
US-09-471-803A-2

Alignment Scores:
Pred. No.: 1,18e-14 Length: 753
Score: 197.00 Matches: 68
Percent Similarity: 46.6% Conservatives: 41
Best Local Similarity: 29.1% Mismatches: 95
Query Match: 15.8% Indels: 30
DB: 3 Gaps: 9

US-10-073-293A-4 (1-245) x US-09-471-803A-2 (1-753)

Qy 17 GluGlyCysLysAspSerLeuProIleValIleSerTyIleProValAlaPheAlaPhe 36
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 100 CAAGGTCTAAACACCTCCCTTGTGCGAGTTTGGCATGTACCGATTTGGTATTGCGTTT 159

Qy 37 GlyLeuAsnAlaThrArgLeuGlyPheSerProLeuGluSerValPhePheSerCysIle 56
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 160 GGTCTCTTGTGTTATTCAATACGGCTACGAATGGTGGCAGCCCACTGTTTTCGGCGCTG 219

Qy 57 IleTyAlaGlyAlaSerGlnPheValIleThrAlaMetLeuAlaAlaGlySerSerLeu 76
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 220 ATTTTTCGGGCTCCACCGAAATGCTGTCATCGCTCTGTTGGGGCGGAGCCCGCTG 279

Qy 77 TrpIleAlaAlaLeuThrValMetAlaMetAspValArgHisValLeuTyGlyProSer 96
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 280 GCGCGCATCGCTCACCACATTTGCTGTAACCTCCGCCACGTTATCTATCGCTTTCA 339

Qy 97 LeuArgSerArgIleIleGlnArgLeuGlnLysSerLysThrAlaLeuTrp-----Ala 114
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 340 TTCCCGCTGTCATGTTGTC-----AAAAACCCCATTCGCGCTTTCTATTTCGTT 387

Qy 115 PheGlyLeuThrAspGluValPheAlaAlaAlaThrAlaLysLeuValArgAsnArg 134
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 388 TTCCGCTTATCCAGCAGCGCTACGACGTACCTGCGCGCAGG-----CCGCGCA 435

Qy 135 ArgTrpSerGluAsnTrpMetIleGlyIleAlaPheSerSerTrpSerSerTrpValPhe 154
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 436 GGTGTTGCGCGGGCGGACTTATCTCAATGCAATAGCGTTTCACTCTACTGGTATTC 495

Qy 155 GlyThrValIleGlyAlaPheSerGlySerGlyLeuLeuGlnGlyTyProAlaValGlu 174
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 496 -----GGCGGTCTTCACCGAGTGGCGATCGCAGAGTTGATTCTCTTTGAAATT 543

```

```

Qy 175 AlaAlaLeuGlyPheMetLeuProAlaLeuPheMetSerPheLeuLeuAlaSerPheGln 194
Db 544 AAGGGCTCGAGTTGGCCCTTGTCTCTTTGTGCACGCTGACTTTGGATTCTCTGCGGA 603
Qy 195 ArgLysGln-----SerLeuCysValThrAlaAlaLeuVal 206
Db 604 ACGAAAAGCAGATCCCTTCTGCTGCTCCAGGTTTGAGCTTCACCATTTGCTTTGTG 663
Qy 207 GlyAlaLeuAlaGlyValThrLeuPheSerIleProValAlaIleLeuAlaGlyIleVal 226
Db 664 ---GTAATTCAGGTCAGGCCCTATTTCG-----GGCTGCTG-----ATCTTC 705
Qy 227 CysGlyCysLeuThrAlaLeuIleGlnAlaPheTrpGlnGly 240
Db 706 TTGGGTCGTGTGACC-----ATCCGGTACTTCTTCTTGGGA 741

RESULT 15
US-10-608-504-2
; Sequence 2, Application US/10608504
; Patent No. 6841360
; GENERAL INFORMATION:
; APPLICANT: KENNERKNECHT, NICOLE
; APPLICANT: SAHM, HERMANN
; APPLICANT: EGGELE, LOTHAR
; APPLICANT: PFEFFERLE, WALTER
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE CODING FOR THE EXPORT OF
; TITLE OF INVENTION: BRANCHED CHAIN AMINO ACIDS, PROCESS FOR THE
; TITLE OF INVENTION: ISOLATION THEREOF AND USE THEREOF
; FILE REFERENCE: 7601/80525
; CURRENT APPLICATION NUMBER: US/10/608,504
; CURRENT FILING DATE: 2003-06-30
; PRIOR APPLICATION NUMBER: US 09/471,803
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: DE 199 51 708.8
; PRIOR FILING DATE: 1999-10-27
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 753
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(753)
; OTHER INFORMATION: brnF
; FEATURE:
; OTHER INFORMATION: ATCC14752
US-10-608-504-2

Alignment Scores:
Pred. No.: 1.18e-14 Length: 753
Score: 197.00 Matches: 68
Percent Similarity: 46.6% Conservative: 41
Best Local Similarity: 29.1% Mismatches: 95
Query Match: 15.6% Indels: 30
DB: 3 Gaps: 9

US-10-073-293A-4 (1-245) x US-10-608-504-2 (1-753)
Qy 17 GluGlyCysLysAspSerLeuProIleValIleSerTyIleProValAlaPheAlaPhe 36
Db 100 CAAGGCTCTAAACCTCCCTTGTCTGCTGACGTTTGGGCGATGTACCCGATTGGTATTGCGTTT 159
Qy 37 GlyLeuAsnAlaThrArgLeuGlyPheSerProLeuGluSerValPhePheSerCysIle 56
Db 160 GGTCTCTGGTTATTCAATACGGCTACGAATGGTGGGCGAGCCCACTGTTTCCGGCCTG 219
Qy 57 IleTyAlaGlyAlaSerGlnPheValIleThrAlaMetLeuAlaAlaGlySerSerLeu 76
Db 220 ATTTTCGGGGCTCCACCGAAATGTGTATCGCCCTCGTTGTGGGCGCAGCGCCCTG 279
Qy 77 TrpIleAlaAlaLeuThrValMetAlaMetAspValArgHisValLeuTyTrpGlyProSer 96

```

```

Db 280 GGGCCATCGCGCTCACCACATTCGTGTAACATTCGCGCAGGTATTTCTATGCGTTC 339
Qy 97 LeuArgSerArgIleIleGlnArgLeuGlnLysSerLysThrAlaLeuTrp-----Ala 114
Db 340 TTCCCGCTGCATGTGGTTC-----AAAAACCCCATTTGCCGTTTCTATTTCGGTT 387
Qy 115 PheGlyLeuThrAspGluValPheAlaAlaAlaThrAlaLysLeuValArgAsnArg 134
Db 388 TTCCCGCTTATCGACGAGGCTACGCACTGCTCCGCGCCAGG-----CCCGCA 435
Qy 135 ArgTrpSerGluAsnTrpMetIleGlyIleAlaPheSerSerTrpSerSerTrpValPhe 154
Db 436 GGCTGGTGGCGGTGGCGACTTATCTCAATGCAATAGCGTTTCACTCCTACTGGGTTTC 495
Qy 155 GlyThrValIleGlyAlaPheSerGlySerGlyLeuLeuGlnGlyTyProAlaValGlu 174
Db 496 -----GGCGGTCTCACCGGAGTGGCGATCGCAGAGTTGATTCTCTTTGAAATT 543
Qy 175 AlaAlaLeuGlyPheMetLeuProAlaLeuPheMetSerPheLeuLeuAlaSerPheGln 194
Db 544 AAGGGCTCGAGTTGCGCCCTTGTCTCTCTTTGTGCACGCTGACTTTGGATTCTGCGGA 603
Qy 195 ArgLysGln-----SerLeuCysValThrAlaAlaLeuVal 206
Db 604 ACGAAAAGCAGATCCCTTCTGCTGCTCGCAGGTTTGAGCTTCACCATTTGCTTTGTG 663
Qy 207 GlyAlaLeuAlaGlyValThrLeuPheSerIleProValAlaIleLeuAlaGlyIleVal 226
Db 664 ---GTAATTCAGGTCAGGCCCTATTTCG-----GGCTGCTG-----ATCTTC 705
Qy 227 CysGlyCysLeuThrAlaLeuIleGlnAlaPheTrpGlnGly 240
Db 706 TTGGGTCGTGTGACC-----ATCCGGTACTTCTTCTTGGGA 741

```

Search completed: February 16, 2006, 19:06:12
Job time : 230.666 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 16, 2006, 17:44:58 ; Search time 946.966 Seconds
(without alignments)
2139.461 Million cell updates/sec

Title: US-10-073-293A-4

Perfect score: 1248
Sequence: 1 MESPTQPAPGSATFMEGCK.....VCGCLTALIQAFWQGPDEL 245

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus_p2n.model -DEV=xlp
-Q=/abs/ABSSWEB.spool/US10073293/runat_15022006_120740_1106/app_query.fasta_1
-DB=Published Applications NA Main -QMT=faetap -SUFFIX=p2n.rnpbm
-MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1
-MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR SCORE=pct
-THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext
-HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abes03p
-USER=US10073293 @CGN_1_1_2064 @runat_15022006_120740_1106 -NCPU=6 -ICPU=3
-NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMBOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA Main.*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1248	100.0	738	9	US-10-073-293A-3
2	426	34.1	1589	3	US-09-764-877-739
3	426	34.1	1589	3	US-09-860-870-69
4	426	34.1	1589	3	US-09-764-875-55
5	426	34.1	1589	6	US-10-227-646-69
6	426	34.1	1589	6	US-10-242-515-739
7	426	34.1	1595	3	US-09-764-877-1016

c	8	426	34.1	1595	3	US-09-764-875-372	Sequence 372, App
c	9	426	34.1	1595	6	US-10-242-515-1016	Sequence 1016, App
c	10	238.5	19.1	12445	3	US-09-070-927A-242	Sequence 242, App
c	11	225.5	18.1	729	7	US-10-724-972A-2949	Sequence 2949, App
c	12	225	18.0	495269	7	US-10-398-221-8	Sequence 8, Appli
c	13	225	18.0	3011208	7	US-10-398-221-2058	Sequence 2058, App
c	14	223	17.9	6157	7	US-10-398-221-3752	Sequence 3752, App
c	15	204.5	16.4	11823	2	US-08-781-986A-136	Sequence 136, App
c	16	204.5	16.4	11823	7	US-10-329-624-136	Sequence 136, App
c	17	197	15.8	753	3	US-09-738-626-289	Sequence 289, App
c	18	197	15.8	753	7	US-10-608-504-2	Sequence 2, Appli
c	19	197	15.8	1271	7	US-10-608-504-1	Sequence 1, Appli
c	20	197	15.8	1271	7	US-10-608-504-6	Sequence 6, Appli
c	21	197	15.8	3309400	3	US-09-738-626-1	Sequence 1, Appli
c	22	195.5	15.7	690	7	US-10-335-977-1445	Sequence 1445, App
c	23	191	15.3	2162598	8	US-10-472-928-4979	Sequence 4979, App
c	24	189	15.1	675	7	US-10-474-776-2	Sequence 2, Appli
c	25	189	15.1	681	9	US-10-501-282-2597	Sequence 2597, App
c	26	189	15.1	681	9	US-10-501-282-2599	Sequence 2599, App
c	27	189	15.1	1754382	9	US-10-501-282-6651	Sequence 6651, App
c	28	184	14.7	4792	2	US-08-961-527-156	Sequence 156, App
c	29	184	14.7	4792	7	US-10-158-844-156	Sequence 156, App
c	30	183	14.7	711	3	US-09-738-626-3408	Sequence 3408, App
c	31	183	14.7	3309400	3	US-09-738-626-1	Sequence 1, Appli
c	32	182.5	14.6	891	6	US-10-156-761-3183	Sequence 3183, App
c	33	182.5	14.6	9025608	6	US-10-156-761-1	Sequence 1, Appli
c	34	178	14.3	654	8	US-10-472-928-95	Sequence 95, Appli
c	35	174.5	14.0	2256846	7	US-10-470-565-1	Sequence 1, Appli
c	36	171	13.7	555	9	US-10-617-320-738	Sequence 738, App
c	37	162	13.0	384	3	US-09-974-300-4863	Sequence 4863, App
c	38	159	12.7	564	3	US-09-974-300-4832	Sequence 4832, App
c	39	153.5	12.3	906	6	US-10-156-761-2485	Sequence 2485, App
c	40	139	11.1	7612	9	US-10-795-159-572	Sequence 572, App
c	41	139	11.1	908766	9	US-10-795-159-685	Sequence 685, App
c	42	136	10.9	1830121	7	US-10-329-670-1	Sequence 1, Appli
c	43	136	10.9	1830121	8	US-10-158-865-1	Sequence 1, Appli
c	44	136	10.9	1830121	9	US-10-981-687-1	Sequence 1, Appli
c	45	126	10.1	19211	9	US-10-915-740A-54	Sequence 54, Appl

ALIGNMENTS

RESULT 1

US-10-073-293A-3
; Sequence 3, Application US/10073293A
; Publication No. US20050239175A1
; GENERAL INFORMATION:
; APPLICANT: TABOLINA, EKATERINA
; APPLICANT: RYBAK, KONSTANTIN
; APPLICANT: KHOURGES, EVGENI
; APPLICANT: VOROSHILOVA, ELVIRA
; APPLICANT: GUSYATINER, MIKHAIL
; TITLE OF INVENTION: METHOD FOR PRODUCING L-AMINO ACID USING BACTERIA BELONGING TO THE
; FILE REFERENCE: 219594USO
; CURRENT APPLICATION NUMBER: US/10/073,293A
; CURRENT FILING DATE: 2002-02-13
; PRIOR APPLICATION NUMBER: RU 2001103865
; PRIOR FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: RU 2001104998
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: RU 2001104999
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: RU 2001117632
; PRIOR FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: RU 2001117633
; PRIOR FILING DATE: 2001-06-28
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 738
; TYPE: DNA
; ORGANISM: Escherichia coli

; FEATURE:

; NAME/KEY: CDS
; LOCATION: (1) ..(735)
; OTHER INFORMATION:

US-10-073-293A-3

Alignment Scores:

Pred. No.: 1.18e-134 Length: 738
Score: 1248.00 Matches: 245
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 9 Gaps: 0

US-10-073-293A-4 (1-245) x US-10-073-293A-3 (1-738)

Qy 1 MetGluSerProThrProGlnProAlaProGlySerAlaThrPheMetGluGlyCysLys 20
Db 1 ATGGAAGCCCTACTCCACAGCCTGCTCTCGTTCGGCGACCTTCATGGAAGGATGCAAA 60
Qy 21 AspSerLeuProIleValIleSerTyrlleProValAlaPheAlaPheGlyLeuAsnAla 40
Db 61 GACAGTTTACCGATGTATTAGTTATATTCGGTGGCCCTTTGGCTCGGFTCTGAATGCG 120
Qy 41 ThrArgLeuGlyPheSerProLeuGluSerValPhePheSerCysIleIleTyrAlaGly 60
Db 121 ACCCGTCTGGGATTTCTCTCTCGAAAGCGTTTTTTCTCTGCATCATTTATGCGAGGC 180
Qy 61 AlaSerGlnPheValIleThrAlaMetIeuAlaAlaGlySerSerLeuTrpIleAlaAla 80
Db 181 GCGAGCCAGTTCGTCAATACCGCATCTGGCAGCCGGAGTAGTTTGTGGATTTGCTGCA 240
Qy 81 LeuThrValMetAlaMetAspValArgHisValLeuTyrGlyProSerLeuArgSerArg 100
Db 241 CTGACCGTCATGCGCAATGAGTGTTCGCCATGTGTGTATGCGCCGTCACCTCGTAGCCGT 300
Qy 101 IleIleGlnArgLeuGlnLysSerLysThrAlaLeuTrpAlaPheGlyLeuThrAspGlu 120
Db 301 ATTATTACGCTCTGCAAAATCGAAACCGCCCTGTGGGGTTTGGCCTGACCGATGAG 360
Qy 121 ValPheAlaAlaThrAlaLysLeuValArgHsnArgArgTrpSerGluAsnTrp 140
Db 361 GTTTTTCGCGCGCAACCGCAAACTGGTAGCGCAATAATCGCCGCTGCGAGCGAGAACTGG 420
Qy 141 MetIleGlyIleAlaPheSerSerTrpSerSerTrpValPheGlyThrValIleGlyAla 160
Db 421 ATGATCGCATTTGCTTTCAGTTTCATGTCATGTCATGTCATGTCATGTCATGTCATGTC 480
Qy 161 PheSerGlySerGlyLeuLeuGlnGlyTyrProAlaValGluAlaAlaLeuGlyPheMet 180
Db 481 TTCTCCGCGACGCGCTGCTGCAAGTTATCCCGCGTTGAAGCTGCATTAAGTTTATG 540
Qy 181 LeuProAlaLeuPheMetSerPheLeuLeuAlaSerPheGlnArgLysGlnSerLeuCys 200
Db 541 CTTCGGGCACCTTTATGAGTTCTCTGCTGCCCTTTTCACGCGCAACAATCTCTTTGC 600
Qy 201 ValThrAlaAlaLeuValGlyAlaLeuAlaGlyValThrLeuPheSerIleProValAla 220
Db 601 GTTACCGCAGCGTTAGTTGGTGGCCCTTCGACGCGTAAGCTATTTCCTATTCCCGTCGCC 660
Qy 221 IleLeuAlaGlyIleValCysGlyCysLeuThrAlaLeuIleGlnAlaPheTrpGlnGly 240
Db 661 ATTCTGGCAGCATTTGCTGTGGCTGCTCTACTGCTTAACTCCAGGCATTTCTGSCAGGA 720
Qy 241 AlaProAspGluLeu 245
Db 721 GCGCCCGATGAGCTA 735

RESULT 2

US-09-764-877-739

; Sequence 739, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:

; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 739
; LENGTH: 1589
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-877-739

Alignment Scores:
Pred. No.: 9.17e-39 Length: 1589
Score: 426.00 Matches: 84
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 34.1% Indels: 0
DB: 3 Gaps: 0

US-10-073-293A-4 (1-245) x US-09-764-877-739 (1-1589)

Qy 1 MetGluSerProThrProGlnProAlaProGlySerAlaThrPheMetGluGlyCysLys 20
Db 1338 ATGGAAGCCCTACTCCACAGCCTGCTCTCGTTCGGCGACCTTCATGGAAGGATGCAAA 1397
Qy 21 AspSerLeuProIleValIleSerTyrlleProValAlaPheAlaPheGlyLeuAsnAla 40
Db 1398 GACAGTTTACCGATGTATTAGTTATATTCGGTGGCCCTTTGGCTCGGFTCTGAATGCG 1457
Qy 41 ThrArgLeuGlyPheSerProLeuGluSerValPhePheSerCysIleIleTyrAlaGly 60
Db 1458 ACCCGTCTGGGATTTCTCTCTCGAAAGCGTTTTTTCTCTGCATCATTTATGCGAGGC 1517
Qy 61 AlaSerGlnPheValIleThrAlaMetIeuAlaAlaGlySerSerLeuTrpIleAlaAla 80
Db 1518 GCGAGCCAGTTCTGTCATTACCGCATGCTGCGACCGGAGTAGTTTGTGGATTTGCTGCA 1577
Qy 81 LeuThrValMet 84
Db 1578 CTGACCGTCATG 1589

RESULT 3

US-09-860-670-69
; Sequence 69, Application US/09860670
; Patent No. US20020165137A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA127P1
; CURRENT APPLICATION NUMBER: US/09/860,670
; CURRENT FILING DATE: 2001-05-21
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 69
; LENGTH: 1589
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-860-670-69

Alignment Scores:
Pred. No.: 9.17e-39 Length: 1589
Score: 426.00 Matches: 84
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 34.1% Indels: 0
DB: 3 Gaps: 0

US-10-073-293A-4 (1-245) x US-09-860-670-69 (1-1589)

QY 1 MetGluSerProThrProGlnProAlaProGlySerAlaThrPheMetGluGlyCysLys 20
Db 1338 ATGGAAGCCCTACTCCACAGCCTGCTCTGTTCCGGACCTTCATGGAAGGATGCAA 1397
QY 21 AspSerLeuProIleValIleSerTyrlleProValAlaPheAlaPheGlyLeuAsnAla 40
Db 1398 GACAGTTTACCGATTGTTATTAGTTATATCCGGTGGCCTTGGCTTCGGTCTGAATGCG 1457
QY 41 ThrArgLeuGlyPheSerProLeuGluSerValPhePheSerCysIleIleTyAlaGly 60
Db 1458 ACCCGTCTGGGATTCCTCTCTCGAAGCGTTTTTTTCTCTGCATCATTTATGCGGC 1517
QY 61 AlaSerGlnPheValIleThrAlaMetLeuAlaGlySerSerLeuTrpIleAlaAla 80
Db 1518 GCGAGCCAGTTTCGTCATTACCGCATGCTGGCAGCCGGAGTAGTTTGTGGATTGCTGCA 1577
QY 81 LeuThrValMet 84
Db 1578 CTGACCGTCATG 1589

RESULT 4

US-09-764-875-55

; Sequence 55, Application US/09764875

; Publication No. US20040018969A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PJ202

; CURRENT APPLICATION NUMBER: US/09/764,875

; CURRENT FILING DATE: 2001-01-17

; Prior application data removed - consult PALM or file wrapper

; NUMBER OF SEQ ID NOS: 1249

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 55

; LENGTH: 1589

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-764-875-55

Alignment Scores:

Pred. No.:	9.17e-39	Length:	1589
Score:	426.00	Matches:	84
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	34.1%	Indels:	0
DB:	3	Gaps:	0

US-10-073-293A-4 (1-245) x US-09-764-875-55 (1-1589)

QY 1 MetGluSerProThrProGlnProAlaProGlySerAlaThrPheMetGluGlyCysLys 20
Db 1338 ATGGAAGCCCTACTCCACAGCCTGCTCTGTTCCGGACCTTCATGGAAGGATGCAA 1397
QY 21 AspSerLeuProIleValIleSerTyrlleProValAlaPheAlaPheGlyLeuAsnAla 40
Db 1398 GACAGTTTACCGATTGTTATTAGTTATATCCGGTGGCCTTGGCTTCGGTCTGAATGCG 1457
QY 41 ThrArgLeuGlyPheSerProLeuGluSerValPhePheSerCysIleIleTyAlaGly 60
Db 1458 ACCCGTCTGGGATTCCTCTCTCGAAGCGTTTTTTTCTCTGCATCATTTATGCGGC 1517
QY 61 AlaSerGlnPheValIleThrAlaMetLeuAlaGlySerSerLeuTrpIleAlaAla 80
Db 1518 GCGAGCCAGTTTCGTCATTACCGCATGCTGGCAGCCGGAGTAGTTTGTGGATTGCTGCA 1577
QY 81 LeuThrValMet 84
Db 1578 CTGACCGTCATG 1589

RESULT 5

US-10-227-646-69

; Sequence 69, Application US/10227646

; Publication No. US20030235829A1

; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA12791
; CURRENT APPLICATION NUMBER: US/10/227,646
; CURRENT FILING DATE: 2002-08-26
; PRIOR APPLICATION NUMBER: US/09/860,670
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: 60/205,515
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 69
; LENGTH: 1589
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-227-646-69

Alignment Scores:

Pred. No.:	9.17e-39	Length:	1589
Score:	426.00	Matches:	84
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	34.1%	Indels:	0
DB:	6	Gaps:	0

US-10-073-293A-4 (1-245) x US-10-227-646-69 (1-1589)

QY 1 MetGluSerProThrProGlnProAlaProGlySerAlaThrPheMetGluGlyCysLys 20
Db 1338 ATGGAAGCCCTACTCCACAGCCTGCTCTGTTCCGGACCTTCATGGAAGGATGCAA 1397
QY 21 AspSerLeuProIleValIleSerTyrlleProValAlaPheAlaPheGlyLeuAsnAla 40
Db 1398 GACAGTTTACCGATTGTTATTAGTTATATCCGGTGGCCTTGGCTTCGGTCTGAATGCG 1457
QY 41 ThrArgLeuGlyPheSerProLeuGluSerValPhePheSerCysIleIleTyAlaGly 60
Db 1458 ACCCGTCTGGGATTCCTCTCTCGAAGCGTTTTTTTCTCTGCATCATTTATGCGGC 1517
QY 61 AlaSerGlnPheValIleThrAlaMetLeuAlaGlySerSerLeuTrpIleAlaAla 80
Db 1518 GCGAGCCAGTTTCGTCATTACCGCATGCTGGCAGCCGGAGTAGTTTGTGGATTGCTGCA 1577
QY 81 LeuThrValMet 84
Db 1578 CTGACCGTCATG 1589

RESULT 6

US-10-242-515-739

; Sequence 739, Application US/10242515

; Publication No. US20040009488A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PC005C1

; CURRENT APPLICATION NUMBER: US/10/242,515

; CURRENT FILING DATE: 2002-09-13

; PRIOR APPLICATION NUMBER: 09/764,877

; PRIOR FILING DATE: 2001-01-17

; PRIOR APPLICATION NUMBER: 60/179,065

; PRIOR FILING DATE: 2000-01-31

; PRIOR APPLICATION NUMBER: 60/180,628

; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: 60/214,886

; PRIOR FILING DATE: 2000-06-28

; PRIOR APPLICATION NUMBER: 60/217,487

; PRIOR FILING DATE: 2000-07-11

; PRIOR APPLICATION NUMBER: 60/225,758

; PRIOR FILING DATE: 2000-08-14

; PRIOR APPLICATION NUMBER: 60/220,963

; PRIOR FILING DATE: 2000-07-26

; PRIOR APPLICATION NUMBER: 60/217,496

```
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 739
; LENGTH: 1589
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-242-515-739

Alignment Scores:
Pred. No.:          9,17e-39      Length:      1589
Score:             426.00      Matches:      84
Percent Similarity: 100.0%      Conservative: 0
Best Local Similarity: 100.0%    Mismatches: 0
Query Match:       34.1%      Indels:      0
DB:                6          Gaps:          0

US-10-073-293A-4 (1-245) x US-10-242-515-739 (1-1589)

Qy      1 MetGluSerProThrProGlnProAlaProGlySerAlaThrPheMetGluGlyCysLys 20
Db      1338 ATGGAAGCCCTACTCCACAGCCTCTCTGTTCCGGGACCTTCATGGAAGGATGCAAA 1397

Qy      21 AspSerLeuProIleValIleSerTyriLeProValAlaPheAlaPheGlyLeuAsnAla 40
Db      1398 GACAGTTACCGATTGTTATTAGTTATATTCGGTGGCGCTTTCGTTCTGATGCG 1457

Qy      41 ThrArgLeuGlyPheSerProLeuGluSerValPhePheSerCysIleIleTyzAlaGly 60
Db      1458 ACCCGTCTGGGATTCTCTCTCTCGAAAGCGTTTTTTTCTCTGCATCATTTATGCGAGC 1517

Qy      61 AlaSerGlnPheValIleThrAlaMetLeuAlaAlaGlySerSerLeuTrpIleAlaAla 80
Db      1518 GCGAGCCAGTTGTCATTACCGCGATGCTGGCAGCGGAGTAGTTTGTGGATTGCTGCA 1577

Qy      81 LeuThrValMet 84
Db      1578 CTGACCGTCATG 1589

RESULT 7
US-09-764-877-1016/c
; Sequence 1016, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1016
; LENGTH: 1595
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (433)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (467)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-877-1016

Alignment Scores:
Pred. No.:          9,21e-39      Length:      1595
Score:             426.00      Matches:      84
Percent Similarity: 100.0%      Conservative: 0
Best Local Similarity: 100.0%    Mismatches: 0
Query Match:       34.1%      Indels:      0
DB:                3          Gaps:          0

US-10-073-293A-4 (1-245) x US-09-764-877-372 (1-1595)

Qy      1 MetGluSerProThrProGlnProAlaProGlySerAlaThrPheMetGluGlyCysLys 20
Db      254 ATGGAAGCCCTACTCCACAGCCTCTCTGTTCCGGGACCTTCATGGAAGGATGCAAA 195

Qy      21 AspSerLeuProIleValIleSerTyriLeProValAlaPheAlaPheGlyLeuAsnAla 40
Db      194 GACAGTTACCGATTGTTATTAGTTATATTCGGTGGCGCTTTCGTTCTGATGCG 135

Qy      41 ThrArgLeuGlyPheSerProLeuGluSerValPhePheSerCysIleIleTyzAlaGly 60
Db      1595 GACAGTTACCGATTGTTATTAGTTATATTCGGTGGCGCTTTCGTTCTGATGCG 135

US-10-073-293A-4 (1-245) x US-09-764-877-1016 (1-1595)

Qy      1 MetGluSerProThrProGlnProAlaProGlySerAlaThrPheMetGluGlyCysLys 20
Db      254 ATGGAAGCCCTACTCCACAGCCTCTCTGTTCCGGGACCTTCATGGAAGGATGCAAA 195

Qy      21 AspSerLeuProIleValIleSerTyriLeProValAlaPheAlaPheGlyLeuAsnAla 40
Db      194 GACAGTTACCGATTGTTATTAGTTATATTCGGTGGCGCTTTCGTTCTGATGCG 135

Qy      41 ThrArgLeuGlyPheSerProLeuGluSerValPhePheSerCysIleIleTyzAlaGly 60
Db      1595 GACAGTTACCGATTGTTATTAGTTATATTCGGTGGCGCTTTCGTTCTGATGCG 135
```

```
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 739
; LENGTH: 1589
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-242-515-739

Alignment Scores:
Pred. No.:          9,17e-39      Length:      1589
Score:             426.00      Matches:      84
Percent Similarity: 100.0%      Conservative: 0
Best Local Similarity: 100.0%    Mismatches: 0
Query Match:       34.1%      Indels:      0
DB:                6          Gaps:          0

US-10-073-293A-4 (1-245) x US-10-242-515-739 (1-1589)

Qy      1 MetGluSerProThrProGlnProAlaProGlySerAlaThrPheMetGluGlyCysLys 20
Db      1338 ATGGAAGCCCTACTCCACAGCCTCTCTGTTCCGGGACCTTCATGGAAGGATGCAAA 1397

Qy      21 AspSerLeuProIleValIleSerTyriLeProValAlaPheAlaPheGlyLeuAsnAla 40
Db      1398 GACAGTTACCGATTGTTATTAGTTATATTCGGTGGCGCTTTCGTTCTGATGCG 1457

Qy      41 ThrArgLeuGlyPheSerProLeuGluSerValPhePheSerCysIleIleTyzAlaGly 60
Db      1458 ACCCGTCTGGGATTCTCTCTCTCGAAAGCGTTTTTTTCTCTGCATCATTTATGCGAGC 1517

Qy      61 AlaSerGlnPheValIleThrAlaMetLeuAlaAlaGlySerSerLeuTrpIleAlaAla 80
Db      1518 GCGAGCCAGTTGTCATTACCGCGATGCTGGCAGCGGAGTAGTTTGTGGATTGCTGCA 1577

Qy      81 LeuThrValMet 84
Db      1578 CTGACCGTCATG 1589

RESULT 7
US-09-764-877-1016/c
; Sequence 1016, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1016
; LENGTH: 1595
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (433)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (467)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-877-1016

Alignment Scores:
Pred. No.:          9,21e-39      Length:      1595
Score:             426.00      Matches:      84
```

Db 134 ACCGCTCGGATTCCTCTCTCGAAGCGTTTTCCTCTCGATCATTTATGCGAGC 75
Qy 61 AlaSerGlnPheValIleThrAlaMetLeuAlaAlaGlySerSerLeuTrpIleAlaAla 80
Db 74 GCGAGCCAGTTTCGTCATTCACCGCATGCTGCGAGCCGGAGTAGTTTGTGGATTGCTGCA 15

Qy 81 LeuThrValMet 84
Db 14 CTGACCGTCATG 3

RESULT 9

US-10-242-515-1016/c
; Sequence 1016, Application US/10242515
; Publication No. US20040009488A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PC005C1

; CURRENT APPLICATION NUMBER: US/10/242,515

; CURRENT FILING DATE: 2002-09-13

; PRIOR APPLICATION NUMBER: 09/764,877

; PRIOR FILING DATE: 2001-01-17

; PRIOR APPLICATION NUMBER: 60/179,065

; PRIOR FILING DATE: 2000-01-31

; PRIOR APPLICATION NUMBER: 60/180,628

; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: 60/214,866

; PRIOR FILING DATE: 2000-06-28

; PRIOR APPLICATION NUMBER: 60/217,487

; PRIOR FILING DATE: 2000-07-11

; PRIOR APPLICATION NUMBER: 60/225,758

; PRIOR FILING DATE: 2000-08-14

; PRIOR APPLICATION NUMBER: 60/220,963

; PRIOR FILING DATE: 2000-07-26

; PRIOR APPLICATION NUMBER: 60/217,496

; PRIOR FILING DATE: 2000-07-11

; PRIOR APPLICATION NUMBER: 60/225,447

; PRIOR FILING DATE: 2000-08-14

; PRIOR APPLICATION NUMBER: 60/218,290

; PRIOR FILING DATE: 2000-07-14

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 4031

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1016

; LENGTH: 1595

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (433)

; OTHER INFORMATION: n equals a,t,g, or c

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (467)

; OTHER INFORMATION: n equals a,t,g, or c

; US-10-242-515-1016

Alignment Scores:

Pred. No.: 9.21e-39 Length: 1595
Score: 426.00 Matches: 84
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 34.1% Indels: 0
DB: 6 Gaps: 0

US-10-073-293A-4 (1-245) x US-10-242-515-1016 (1-1595)

Qy 1 MetGluSerProThrGlnProAlaProGlySerAlaThrPheMetGluGlyCysLys 20
Db 254 ATGGAAGCCCTACTCCACACCCCTCTCTGTTGGCGACCTTCATGGAGGATGCAAA 195

Qy 21 AspSerLeuProIleValIleSerTyrIleProValAlaPheAlaPheGlyLeuAsnAla 40
|||||

Db 194 GACAGTTTACCGATTTGTTATTATTATATATTCGGTGGCCTTTGGGTTGGTCTGAATGCG 135
Qy 41 ThrArgLeuGlyPheSerProLeuGluSerValPhePheSerCysIleIleIleTyrAlaGly 60
Db 134 ACCGCTCGGATTCCTCTCTCGAAGCGTTTTCCTCTCGATCATTTATGCGAGC 75
Qy 61 AlaSerGlnPheValIleThrAlaMetLeuAlaAlaGlySerSerLeuTrpIleAlaAla 80
Db 74 GCGAGCCAGTTTCGTCATTCACCGCATGCTGCGAGCCGGAGTAGTTTGTGGATTGCTGCA 15
Qy 81 LeuThrValMet 84
Db 14 CTGACCGTCATG 3

RESULT 10

US-09-070-927A-242/c

; Sequence 242, Application US/09070927A

; Patent No. US20020120116A1

; GENERAL INFORMATION:

; APPLICANT: Charles A. Kunsch

; PATENT ATTORNEY: Steven Barash

; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides

; NUMBER OF SEQUENCES: 982

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/070,927A

; FILING DATE: 04-May-2000

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/046,655

; FILING DATE: 1997-05-16

; APPLICATION NUMBER: 60/044,031

; FILING DATE: 1997-05-06

; APPLICATION NUMBER: 60/066,009

; FILING DATE: 1997-11-14

; ATTORNEY/AGENT INFORMATION:

; NAME: Kenley K. Hoover

; REGISTRATION NUMBER: 40,302

; REFERENCE/DOCKET NUMBER: PB369

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (301) 309-8504

; TELEFAX: (301) 309-8512

; INFORMATION FOR SEQ ID NO: 242:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 12445 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; SEQUENCE DESCRIPTION: SEQ ID NO: 242:

US-09-070-927A-242

Alignment Scores:

Pred. No.: 9.19e-16 Length: 12445
Score: 238.50 Matches: 71
Percent Similarity: 46.4% Conservative: 33
Best Local Similarity: 31.7% Mismatches: 99
Query Match: 19.1% Indels: 21
DB: 3 Gaps: 7

US-10-073-293A-4 (1-245) x US-09-070-927A-242 (1-12445)

Qy 18 GlyCysLysAspSerLeuProileValileSerTyrlleProValAlaPheAlaPheGly 37
 Db 5700 GGCATTCGTGAAGCCCTTCCACAGGTTTGGCTATATTTGGTATGGCATTCGTTTGGT 5641
 Qy 38 LeuAenAlaThrArgLeuGlyPheSerProLeuGluSerValPhePheSerCysIlelle 57
 Db 5640 ATCGTAGGACAAAGCAGCAGGTTTTCCTCTTATTTAGTGGCATGTCGTTTATTTATT 5581
 Qy 58 TyrAlaGlyAlaSerGlnPheValileThrAlaMetLeuAlaAlaGlySerSerLeuTrp 77
 Db 5580 TTTGAGGATCTGCTAGTTGTAAACAGTTAGTATGTTGACTGGCGGAGCCCTTCTTA 5521
 Qy 78 IleAlaAlaLeuThrValMetAlaMetAspValArgHisValLeuTyTrpSerSerLeu 97
 Db 5520 TCTATTGTTAGCGACCTTCTTAGTAATGCGGTATGATCTTATGGGATGACCAT 5461
 Qy 98 ArgSerArgIle---IleGlnArgLeuGlnIysSerIysThrAlaLeuTrpAlaPheGly 116
 Db 5460 GCGCCTTACTTTAAAGCCGAATCTCTCGTAAATAAT-----CTCTGG---TTGGGA 5413
 Qy 117 -----LeuThrAspGluValPheAlaAlaAlaThrAlaAlaLeuValArgAsnArg 134
 Db 5412 ACCTTACTGACGAGTGAAGTTTTCGACTAGGAATGAATAAATTAACACCAACCAAGGA 5353
 Qy 135 ArgTrpSerGluAsnTrpMetIleGlyIleAlaPheSerSerTrpSerTrpValPhe 154
 Db 5352 CGTTTAAAGTTTGAAGTGTAAATGCTGCCAATTTGATTTCTATGGCAGCTGGGTTT 5293
 Qy 155 GlyThrValIleGlyAlaPheSerGlySerGlyLeuLeuGlnGlyTyTrpProAlaValGlu 174
 Db 5292 TCGACGATCATCGGTGCTATCTAGGGGC-----TTTATGCAACCCCT 5248
 Qy 175 AlaAlaLeuGly-----PheMetLeuProAlaLeuPheMetSerPheLeuLeuAlaSer 192
 Db 5247 CAAGCATTAGGCTTTCGAGTTCGCGTGTGCTGCTATGTTTCATTTGGTTTACTCTATCTGCAG 5188
 Qy 193 PheGlnArgLysGlnSerLeuCysValThrAlaAlaLeuValGlyAlaLeuAlaGlyVal 212
 Db 5187 ATAATTAGTATCGTTCATGNAATCGCTTACACCTCATCGTCTCATTTACTTTT 5128
 Qy 213 ThrLeuPheSerIle-----ProValAlaIleLeuAlaGlyIle 225
 Db 5127 GGTTTGATGCTATTGGTCTAATTTTATCCAGTAATTTAAATGTCYTAGTTGKCACG 5068
 Qy 226 ValCysGlyCys 229
 Db 5067 TTAATCGGGTGT 5056

RESULT 11

US-10-724-972A-2949
 ; Sequence 2949, Application US/10724972A
 ; Publication No. US20040147734A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Doucette-Stamm, Lynn
 ; APPLICANT: Bush, David
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: PATH03-16
 ; CURRENT APPLICATION NUMBER: US/10/724,972A
 ; PRIOR FILING DATE: 2003-12-01
 ; PRIOR APPLICATION NUMBER: 09/450,969
 ; PRIOR FILING DATE: 1999-11-29
 ; PRIOR APPLICATION NUMBER: 09/134,001
 ; PRIOR FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 7544
 ; SEQ ID NO 2949
 ; LENGTH: 729
 ; TYPE: DNA
 ; ORGANISM: S.epidermidis

US-10-724-972A-2949
 Alignment Scores:
 Pred. No.: 5.92e-16 Length: 729
 Score: 225.50 Matches: 66
 Percent Similarity: 45.5% Conservative: 39
 Best Local Similarity: 28.6% Mismatches: 89
 Query Match: 18.1% Indels: 37
 Dbs: 7 Gaps: 7

US-10-073-293A-4 (1-245) x US-10-724-972A-2949 (1-729)

Qy 14 ThrPheMetGluGlyCysLysAspSerLeuProileValileSerTyrlleProValAla 33
 Db 49 ACGTTTAAACAAGGTGTGAAGAGTGTATCCCACGTTACTTGGTTATGCTGGGTAGGA 108
 Qy 34 PheAlaPheGlyLeuAenAlaThrArgLeuGlyPheSerProLeuGluSerValPhePhe 53
 Db 109 CTATCGTTTGAATTTGGCAGCCCTCCCAAAATTTTCAGTGTGTTTAGAAAATTTATTTATTG 168
 Qy 54 SerCysIleIleTyrlleGlyAlaSerGlnPheValileThrAlaMetLeuAlaAlaGly 73
 Db 169 TGTCTGATATTATTAATGCTGGTGCAGCCAAATTTATTTATTTGTACATGATGTCAGGC 228
 Qy 74 SerSerLeuTrpIleAlaAlaLeuThrValMetAlaMetAspValArgHisValLeuTrp 93
 Db 229 ACCCTATTTCGCAATTGCTTACACACTTATCGTTAACTCTCGAATGTTCTTATTA 288
 Qy 94 GlyProSerLeuArgSerArgIleIleGlnArgLeuGlnIysSerIysThrAlaLeuTrp 113
 Db 289 AGTATGACTTTAGCACCCCAATTTATAAGCAATATGATTTTGGATAGGTAGGCTTGGGA 348
 Qy 114 AlaPheGlyLeuThrAspGluValPheAlaAlaThrAlaLysLeuValArgAsnArg 133
 Db 349 ACGTTA---TTAACAGATGAAACCTTTTGGAGTTGCTATAACGCCATATGTCAA---GGT 402
 Qy 134 ArgArgTrpSerGluAsnTrpMetIleGlyIleAlaPheSerSerTrpSerTrpVal 153
 Db 403 GAAAAAATTAAAGATCATGCTACACGACTAAATATTACTGCTTACTTATTTGGACT 462
 Qy 154 PheGlyThrValIleGlyAlaPheSerGlySerGlyLeuLeuGlnGlyTyTrpProAlaVal 173
 Db 463 GTTTTCATGTGTAATCGGTGCCATTTTCGGAGAG-----TATATTTCAAAT 507
 Qy 174 GluAlaAlaLeuGly-----PheMetLeuProAlaLeuPheMetSerPheLeuLeuAla 191
 Db 508 CCTGATCGCTTGGCCTAGACTTTTGCCATTACCGCAATGTTTATTTTATGATATATCT 567
 Qy 192 SerPheGln-----ArgLysGlnSerLeu-----CysVal 201
 Db 568 CAATTCGAGGGATTAGAAATCAGATTGAGATATATATTGTACTCATTTGATGTGTG 627
 Qy 202 -----ThrAla 203
 Db 628 ATGTGATGATGCTTCTTAAAGTTCAATTTCTACCTTCATACGTAGCAAAATTTTAATAGCC 687
 Qy 204 AlaLeuValGlyAlaLeuAlaGlyValThrLeu 214
 Db 688 GCAATTTGCTGCAATTTGTTAGGGGTGTTGATG 720

RESULT 12

US-10-398-221-8/c
 ; Sequence 8, Application US/10398221
 ; Publication No. US20040018514A1
 ; GENERAL INFORMATION:
 ; APPLICANT: KUNST, Frederik
 ; APPLICANT: GLASER, Philippe
 ; TITLE OF INVENTION: Listeria innocua, genome and applications
 ; FILE REFERENCE: 344 702 - US
 ; CURRENT APPLICATION NUMBER: US/10/398,221
 ; PRIOR FILING DATE: 2003-03-27
 ; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
 ; PRIOR FILING DATE: 2001-10-04

;; PRIOR APPLICATION NUMBER: FR 00/12 697
;; PRIOR FILING DATE: 2000-10-04
;; NUMBER OF SEQ ID NOS: 4025
;; SOFTWARE: PatentIn version 3.0
;; SEQ ID NO 8
;; LENGTH: 495269
;; TYPE: DNA
;; ORGANISM: Listeria innocua
;; FEATURE:
;; NAME/KEY: misc feature
;; LOCATION: (1)..(end)
;; OTHER INFORMATION: n can be any nucleotide: a, g, c or t/u
US-10-398-221-8

Alignment Scores:
Pred. No.: 5,46e-12 Length: 495269
Score: 225.00 Matches: 67
Percent Similarity: 46.2% Conservative: 41
Best Local Similarity: 28.6% Mismatches: 112
Query Match: 18.0% Indels: 14
DB: 7 Gaps: 7

US-10-073-293A-4 (1-245) x US-10-398-221-8 (1-495269)

Qy	14	ThrPheMetGluGlyCysLysAspSerLeuProIleValIleSerTyrlleProValAla	33
Db	275958	AGTTTTATGATGCTGTGAAGCTTGTTCGCAACTGTACTAGTTATGCTGGTATTGGA	275899
Qy	34	PheAlaPheGlyLeuAenAlaThrArgLeuGlyPheSerProLeuGluSerValPhePhe	53
Db	275898	ATAGCTGCTGTGTGTGGAGGGCTTCTCACTTAAGCGTCTTAGAAGTAACGTTACTC	275839
Qy	54	SerCysIleIleTyAlaGlyAlaSerGlnPheValIleThrAlaMetLeuAlaAaGly	73
Db	275838	GCTATTATCGTCTATGCTGGAGCTGCTCAGTTATTATTCGGATTATTATTATTACAA	275779
Qy	74	SerSerLeuTrpIleAlaAlaLeuThrValMetAlaMetAspValArgHisValLeuTy	93
Db	275778	AGCCCTATTCAGCTATTATTATTTTACTACTTTTTTAATAAATTAATTCACGCCATTT	275719
Qy	94	GlyProSerLeuArgSerArgIleGlnArgLeuGlnLysSerTyrlleThrAlaLeuTrp	113
Db	275718	AGTATGCGCGAAGCTCTCATTTTAAAGAGTATTCCTTTTGAATAACATTTGGGATTGGG	275659
Qy	114	AlaPheGlyLeuThrAspGluValPheAlaAlaAlaThrAlaLysLeuValArgAsn	133
Db	275658	GCGCTT---CTGACGACGAAACATTTGGCGTTTCAATGAATCAA---ATTGGCAATAA	275605
Qy	134	ArgArgTrpSerGluAenTrpMetIleGlyIleAlaPheSerSerTrpSerTrpVal	153
Db	275604	AAACCAAGTTAGCGCAAAATGGATGCACGGAATTAACGTAACAGCTACATAGCTGGATT	275545
Qy	154	PheGlyThrValIleGlyAlaPheSerGlySerGlyLeuGlnGlyTyrlleProAlaVal	173
Db	275544	GCGCTTGTATTATTAGTTCTTTTCATTTGGGAATTTGGTTG-----CCAAATCCA	275497
Qy	174	GluAla---AlaLeuGlyPheMetLeuProAlaLeuPheMetSerPheLeu---LeuAla	191
Db	275496	GAACAGTTTGGTCTTGATTTTGGCTTATCAGCAATGTTATTCGGCTGTCTATTATCAA	275437
Qy	192	SerPheGlnArgLysGlnSerLeuCyseValThrAlaAlaLeuValGlyAlaLeuAaGly	211
Db	275436	GTCTTAGCGATAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG	275377
Qy	212	ValThrLeu-----PheSerIlePro---ValAlaIleLeuAlaGlyIle	225
Db	275376	ATTATTATTAATCGGTTTATCGGTTTATGACACCAAGTATGACCAATTTAGCGCTTACC	275317
Qy	226	ValCysGlyCysLeuThrAlaLeuIleGlnAlaPheTrpGln	239
Db	275316	TTACTTGTGTTTGTGATGGAGTGATAGTTGAAAAATGGCAA	275275

RESULT 13

US-10-398-221-2058
;; Sequence 2058, Application US/10398221
;; Publication No. US20040018514A1
;; GENERAL INFORMATION:
;; APPLICANT: KUNST, Frederik
;; APPLICANT: GLASER, Philippe
;; TITLE OF INVENTION: Listeria innocua, genome and applications
;; FILE REFERENCE: 344 702 - US
;; CURRENT APPLICATION NUMBER: US/10/398,221
;; PRIOR FILING DATE: 2003-03-27
;; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
;; PRIOR FILING DATE: 2001-10-04
;; PRIOR APPLICATION NUMBER: FR 00/12 697
;; PRIOR FILING DATE: 2000-10-04
;; NUMBER OF SEQ ID NOS: 4025
;; SOFTWARE: PatentIn version 3.0
;; SEQ ID NO 2058
;; LENGTH: 3011208
;; TYPE: DNA
;; ORGANISM: Listeria innocua
US-10-398-221-2058

Alignment Scores:
Pred. No.: 6,59e-11 Length: 3011208
Score: 225.00 Matches: 67
Percent Similarity: 46.2% Conservative: 41
Best Local Similarity: 28.6% Mismatches: 112
Query Match: 18.0% Indels: 14
DB: 7 Gaps: 7

US-10-073-293A-4 (1-245) x US-10-398-221-2058 (1-3011208)

Qy	14	ThrPheMetGluGlyCysLysAspSerLeuProIleValIleSerTyrlleProValAla	33
Db	1482994	AGTTTTATGATGCTGTGAAGCTTGTTCGCAACTGTACTAGTTATGCTGGTATTGGA	1483053
Qy	34	PheAlaPheGlyLeuAenAlaThrArgLeuGlyPheSerProLeuGluSerValPhePhe	53
Db	1483054	ATAGCTGCTGTGTGTGGAGGGCTTCTCACTTAAGCGTCTTAGAAGTAACGTTACTC	1483113
Qy	54	SerCysIleIleTyAlaGlyAlaSerGlnPheValIleThrAlaMetLeuAlaAaGly	73
Db	1483114	GCTATTATCGTCTATGCTGGAGCTGCTCAGTTATTATTTTGGGATTATTATTATTACAA	1483173
Qy	74	SerSerLeuTrpIleAlaAlaLeuThrValMetAlaMetAspValArgHisValLeuTy	93
Db	1483174	AGCCCTATTCAGCTATTATTATTTTACTACTTTTTTAATAAATTAATTCACGCCATTT	1483233
Qy	94	GlyProSerLeuArgSerArgIleGlnArgLeuGlnLysSerTyrlleThrAlaLeuTrp	113
Db	1483234	AGTATGCGCGAAGCTCTCATTTTAAAGAGTATTCCTTTTGAATAACATTTGGGATTGGG	1483293
Qy	114	AlaPheGlyLeuThrAspGluValPheAlaAlaAlaThrAlaLysLeuValArgAsn	133
Db	1483294	GCGCTT---CTGACGACGAAACATTTGGCGTTTCAATGAATCAA---ATTGGCAATAA	1483347
Qy	134	ArgArgTrpSerGluAenTrpMetIleGlyIleAlaPheSerSerTrpSerTrpVal	153
Db	1483348	AAACCAAGTTAGCGCAAAATGGATGCACGGAATTAACGTAACAGCTACATAGCTGGATT	1483407
Qy	154	PheGlyThrValIleGlyAlaPheSerGlySerGlyLeuGlnGlyTyrlleProAlaVal	173
Db	1483408	GCGCTTGTATTATTAGTTCTTTTCATTTGGGAATTTGGTTG-----CCAAATCCA	1483455
Qy	174	GluAla---AlaLeuGlyPheMetLeuProAlaLeuPheMetSerPheLeu---LeuAla	191
Db	1483456	GAACAGTTTGGTCTTGATTTTGGCTTATCAGCAATGTTATTCGGCTGTCTATTATCAA	1483515
Qy	192	SerPheGlnArgLysGlnSerLeuCyseValThrAlaAlaLeuValGlyAlaLeuAaGly	211
Db	1483516	GTCTTAGCGATAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG	1483575
Qy	212	ValThrLeu-----PheSerIlePro---ValAlaIleLeuAlaGlyIle	225

```
Db 1483576 ATTTTATTAATCGTGTATGCGTTTATGACACCAAGATTAGCAATTTAGCGGTACC 1483635
Qy 226 ValCysGlyCysLeuThrAlaLeuLeuGlnAlaPheTrpGln 239
Db 1483636 TTACTTGGTTGTTTGAATGGAGTGATAGTTGAAAAATGGCAA 1483677

RESULT 14
US-10-398-221-3752
; Sequence 3752, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederick
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3752
; LENGTH: 6157
; TYPE: DNA
; ORGANISM: Listeria monocytogenes 4b
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(end)
; OTHER INFORMATION: n can be any nucleotide: a,g,c or t/u
US-10-398-221-3752

Alignment Scores:
Pred. No.: 2,19e-14 Length: 6157
Score: 223.00 Matches: 70
Percent Similarity: 44.6% Conservative: 38
Best Local Similarity: 28.9% Mismatches: 110
Query Match: 17.9% Indels: 24
DB: 7 Gaps: 8

US-10-073-293A-4 (1-245) x US-10-398-221-3752 (1-6157)
Qy 14 ThrPheMetGluGlyCysLeuAspSerLeuProIleValIleSerTyIleProValAla 33
Db 4269 AGTTTTATGATGGTGTAGAGCGCTCTCTACCGTCTTGGTGTATGCTGGGATTGGT 4328
Qy 34 PheAlaPheGlyLeuAsnAlaThrArgLeuGlyPheSerProLeuGluSerValPhePhe 53
Db 4329 ATCGCTGCAGGATAGTAGGAAAAGCATCCCATTTAAGCCCTTTTAGAAGTGACGCTACTT 4388
Qy 54 SerCysIleIleTyAlaGlyAlaSerGlnPheValIleThrAlaMetLeuAlaGly 73
Db 4389 CGGATTATCGTTATGAGGTGGCGGCAATTTATTTCTGTTTGTATTACTACAA 4448
Qy 74 SerSerLeuTrpIleAlaAlaLeuThrValMetAlaMetAspValArgHisValLeuTy 93
Db 4449 AGTCGATATCAGCAATCATTTTCTACTTCTTTTAAATTAATCAAGGCAATTTTAAATG 4508
Qy 94 GlyProSerLeuArgSerArgIleIleGlnArgLeuGlnLysSerIleThrAlaLeuTrp 113
Db 4509 AGT-----ATGCGGAAGCTCCTCATTTTAAAGAGTATTTCTTTGGG 4550
Qy 114 -----AlaPheGly-----LeuThrAspGluValPheAlaAlaAlaThrAlaLys 128
Db 4551 AATAATATTGGATTGGTGGCCCTTCTGACGATGAAAGCTTTGGGTTTCGATGATCAG 4610
Qy 129 LeuValArgAsnAsnArgArgTrpSerGluAsnTrpMetIleGlyIleAlaPheSerSer 148
Db 4611 ---ATTGGTAATAAAAAACCTGTTAGTCTAAATGGATGCACGGAATAAATGTAATGCC 4667
Qy 149 TrpSerSerTrpValPheGlyThrValIleGlyAlaPheSerGlySerGlyLeuLeuGln 168
```

```
Db 4668 TACTTGGCATGATTGTGCTTGTATCGTTGGGCAATTTATTGGCAACTGGCTC----- 4721
Qy 169 GlyTyTrpProAlaValGluAla---AlaLeuGlyPheMetLeuProAlaLeuPheMetSer 187
Db 4722 -----CCAAATCCAGAACAAATTTGGTTTACAGCTTTCGCTGTGCTGCAATGTTTATCGGT 4775
Qy 188 PheLeu---LeuAlaSerPheGlnArgLysGlnSerLeuCysValThrAlaAlaLeuVal 206
Db 4776 TTGCTATATTATCAAGTAGTAGTGATATAAAGCAAAATAAGTAGAGTTTGTTCGTC 4835
Qy 207 GlyAlaLeuAlaGlyValThrLeu-----PheSerIleProValAla 220
Db 4836 ATGATACTAGTCGCTGATTTTAAATCTCTTTATGCGTGTGATGACGCCAGAACTAGCT 4895
Qy 221 IleLeuAlaGlyIleValCysGlyCysLeuThrAlaLeuIleGlnAlaPheTrpGlnGly 240
Db 4896 ATTTTAACTGCACCTTACTCGGATGTTTGATGGGAGTGATTATAGAGATGCGGTAA 4955
Qy 241 AlaPro 242
Db 4956 GTTCCT 4961

RESULT 15
US-08-781-986A-136/c
; Sequence 136, Application US/08781986A
; Publication No. US20030054436A1
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,986A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 136:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11823 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-781-986A-136

Alignment Scores:
Pred. No.: 7,58e-12 Length: 11823
Score: 204.50 Matches: 65
Percent Similarity: 47.0% Conservative: 45
Best Local Similarity: 27.8% Mismatches: 97
Query Match: 16.4% Indels: 27
DB: 2 Gaps: 10
```


This Page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 16, 2006, 17:47:13 ; Search time 655.857 Seconds
(without alignments)
792.964 Million cell updates/sec

Title: US-10-073-293A-4
Perfect score: 1248
Sequence: 1 MESPTQPAPGSAFMFGCK.....VCGCLTALIOAFWQGAPDEL 245

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Xgapop 10.0, Xgapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 7204252 seqs, 1061369211 residues

Total number of hits satisfying chosen parameters: 14408504

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/abs/ABSSWEB.spool/US10073293/runat_15022006_120743_1175/app_query.fasta.1
-DB=Published Applications NA New -QFMT=fastp -SURFIX=p2n.rnpbn -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -HOST=abs03p
-USER=US10073293 @CGN_1_1579 @runat_15022006_120743_1175 -NCPU=6 -ICPU=3
-NO MMAP -NEG SCORES=0 -WAIT -DSPBIOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA New:

1: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
2: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US05_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	229.5	18.4	693	8	US-10-793-626-267
2	229.5	18.4	3058	8	US-10-793-626-3585
3	229.5	18.4	3077	8	US-10-793-626-3594
4	108	8.7	4381	12	US-11-136-527-2390

5	105	8.4	1035	11	US-11-082-389-13	Sequence 13, Appl
6	93	7.5	2748	8	US-10-467-657-5231	Sequence 5231, Ap
7	93	7.5	2763	8	US-10-467-657-7353	Sequence 7353, Ap
8	91.5	7.3	1299	8	US-10-873-528-229	Sequence 229, App
9	91.5	7.3	1947	8	US-10-793-626-1059	Sequence 1059, Ap
10	91.5	7.3	2397	8	US-10-793-626-3342	Sequence 3342, Ap
11	91.5	7.3	3298	8	US-10-793-626-3792	Sequence 3792, Ap
12	91.5	7.3	3332	12	US-11-136-527-74	Sequence 74, Appl
13	90.5	7.3	1419	8	US-10-467-657-1873	Sequence 1873, Ap
14	90	7.2	110711	8	US-10-995-561-13264	Sequence 13264, A
15	89.5	7.2	168516	12	US-11-121-086-3	Sequence 3, Appli
16	87.5	7.0	1706	12	US-11-012-668-3	Sequence 3, Appli
17	87.5	7.0	3639	8	US-10-793-626-3393	Sequence 3393, Ap
18	87	7.0	1416	8	US-10-467-657-2267	Sequence 2267, Ap
19	87	7.0	168753	12	US-11-181-234-1	Sequence 1, Appli
20	86.5	6.9	4041	8	US-10-793-626-3573	Sequence 3573, Ap
21	86.5	6.9	67858	8	US-10-995-561-13484	Sequence 13484, A
22	86.5	6.9	124972	12	US-11-121-086-100	Sequence 100, App
23	86	6.9	903	12	US-11-098-686-9238	Sequence 9238, Ap
24	86	6.9	2469	8	US-10-467-657-2525	Sequence 2525, Ap
25	86	6.9	33931	12	US-11-124-367A-5006	Sequence 5006, Ap
26	86	6.9	1457619	12	US-11-098-686-8739	Sequence 8739, Ap
27	85.5	6.9	1362	8	US-10-793-626-3259	Sequence 3259, Ap
28	85.5	6.9	3181	8	US-10-793-626-3673	Sequence 3673, Ap
29	85.5	6.9	88421	12	US-11-205-109-1	Sequence 1, Appli
30	85	6.8	1072	6	US-09-925-065A-87567	Sequence 87567, A
31	84.5	6.8	2089	12	US-11-210-316-25	Sequence 25, Appl
32	84	6.7	2214	11	US-11-082-389-361	Sequence 361, App
33	83.5	6.7	1524	8	US-10-467-657-7825	Sequence 7825, Ap
34	83.5	6.7	3158	8	US-10-821-234-175	Sequence 175, App
35	83	6.7	697	6	US-09-925-065A-876781	Sequence 876781, A
36	83	6.7	1452	8	US-10-467-657-3819	Sequence 3819, Ap
37	83	6.7	1617	8	US-10-467-657-3145	Sequence 3145, Ap
38	83	6.7	1910	12	US-11-136-527-1918	Sequence 1918, Ap
39	82.5	6.6	1425	12	US-11-074-176-335	Sequence 335, App
40	82.5	6.6	1464	12	US-11-074-176-125	Sequence 125, App
41	82.5	6.6	14896	9	US-11-245-147-243	Sequence 243, App
42	82.5	6.6	14896	12	US-11-000-688-946	Sequence 946, App
43	82.5	6.6	14896	12	US-11-076-427A-31	Sequence 31, Appl
44	82	6.6	1434	6	US-09-925-065A-705332	Sequence 705332, A
45	82	6.6	2079	8	US-10-467-657-6175	Sequence 6175, Ap

ALIGNMENTS

RESULT 1
US-10-793-626-267
; Sequence 267, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10793,626
; PRIOR FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patentin ver. 2.1
; SEQ ID NO 267
; LENGTH: 693
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-267

Alignment Scores:
Pred. No.: 5.63e-14
Score: 229.50
Percent Similarity: 49.3%
Best Local Similarity: 28.6%
Length: 693
Matches: 65
Conservative: 47
Mismatch: 102


```

; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3594
; LENGTH: 3077
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-3594

Alignment Scores:
Pred. No.: 3,21e-13 Length: 3077
Score: 229.50 Matches: 65
Percent Similarity: 49.3% Conservative: 47
Best Local Similarity: 28.6% Mismatches: 102
Query Match: 18.4% Indels: 13
DB: 8 Gaps: 6

US-10-073-293A-4 (1-245) x US-10-793-626-3594 (1-3077)

Qy 14 ThrPheMetGluGlyCysLysAspSerLeuProLleValIleSerTyrIleProValAla 33
Db 2396 ACGTTTAAACAGGTGTGAAGATGTATTCACAGTTACTTGGTTATGCGAGGTAGGA 2455

Qy 34 PheAlaPheGlyLeuAsnAlaThrArgLeuGlyPheSerProLeuGluSerValPhePhe 53
Db 2456 CTATCGTTTGGAAATGTGGCAGTCTCCCAAAATTTCAAGTGTTTAGAAATTTATTTATTG 2515

Qy 54 SerCysIleIleTyrAlaGlyAlaSerGlnPheValIleThrAlaMetLeuAlaAlaGly 73
Db 2516 TGTCTGAATTTATGCTGGTGCAGCTCAATTTATTTATTTGTACATAGTAGTGCAGGC 2575

Qy 74 SerSerLeuTyrIleAlaAlaLeuThrValMetAlaMetAspValArgHisValLeuTyr 93
Db 2576 ACCCTATTCTGCAATGTGCTTACAAATACCTTATCTGTAATCTCGAATGTTCTTATTA 2635

Qy 94 GlyProSerLeuArgSerArgIleIleGlnArgLeuGlnLysSerLysThrAlaLeuTyr 113
Db 2636 AGTATGACTTTAGCACCTAATTAAGCAATATGGAATTTTGGAAATAGGGTAGGGCTTGA 2695

Qy 114 AlaPheGlyLeuThrAspGluValPheAlaAlaAlaThrAlaLysLeuValArgAsn 133
Db 2696 ACGTTA---TTAACAGATGAACATTTTGGCGTTCCTATAACACCATATGTTAA---GGT 2749

Qy 134 ArgArgTyrSerGluAsnTyrMetIleGlyIleAlaPheSerSerTyrSerTyrVal 153
Db 2750 GAAAAAATTAACGATCGATGCTACAGGACTAAATATTAATCTGTTACTTATTTTGGACT 2809

Qy 154 PheGlyThrValIleGlyAlaPheSerGlySerGlyLeuLeuGlnGlyTyrProAlaVal 173
Db 2810 GTTTCTGTGTAAATCGGTGCCATTTTCGGAGAG-----TATATTTCAAT 2854

Qy 174 GluAlaAlaLeuGly-----PheMetLeuProAlaLeuPheMetSerPheLeuLeuAla 191
Db 2855 CCTGATCGCTCGGCCTAGACTTTTGCCAATTAACCGCAATGTTATTTTATGATATCT 2914

Qy 192 SerPheGln-----ArgLysGlnSerLeuCysValThrAlaAlaLeuValGlyAlaLeu 209
Db 2915 CAATTTGAAGGGATTAGAAATCAAGATGGAATATATATGATGATGATGATGATGATG 2974

Qy 210 AlaGlyValThrLeuPhe-----SerIleProValAlaIleLeuAlaGlyIleValCys 227
Db 2975 ATTGTGATGATGCTTCTTCTAAGTTCAATTTCTACCTTCATACCTAGCAATTTAATAGCC 3034

Qy 228 GlyCysLeuThrAlaLeuIle 234
Db 3035 GCAATTGTGTGCATTGTGA 3055

RESULT 4
US-11-136-527-2390
; Sequence 2390, Application US/11136527
; Publication No. US2005028750A1

```

```

; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2390
; LENGTH: 4381
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-2390

Alignment Scores:
Pred. No.: 1.44 Length: 4381
Score: 108.00 Matches: 65
Percent Similarity: 34.7% Conservative: 34
Best Local Similarity: 22.8% Mismatches: 82
Query Match: 8.7% Indels: 104
DB: 12 Gaps: 14

US-10-073-293A-4 (1-245) x US-11-136-527-2390 (1-4381)

Qy 1 MetClnSerProThrProGlnProAlaProGlySerAlaThr----- 14
Db 101 ATGAGGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 160

Qy 15 -----PheMetGluGlyCysLysAspSerLeuProLleValIleSerTyr 29
Db 161 AGCTGCGCAACGCTACGTGAGAGAGCTGC----- 190

Qy 30 IleProValAlaPheAlaPheGlyLeuAsnAlaThrArgLeuGlyPheSerProLeuGlu 49
Db 191 -----GTGAGCGCGCTGCGCGCAGGACCTCCCGCGC 220

Qy 50 SerValPhePhe-----SerCysIleIleTyr 58
Db 221 AGCTTCTTCTTCCACCTCTCGCTGCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAG 280

Qy 59 AlaGlyAla-----SerGlnPheValIleThrAlaMetLeuAla 71
Db 281 GCCGGGCGACGCTCTCGCTGCGCGCTCTGCGCGCTTGTCTGCGCGCGCTGCGCGCGCG 340

Qy 72 AlaGlySerSerLeuTyrIleAlaAlaLeuThrValMetAlaMetAspValArgHisVal 91
Db 341 GCGGGGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 388

Qy 92 LeuTyrGlyProSerLeuArgSerArgIleIleGlnArgLeuGlnLysSerLysThrAla 111
Db 389 CTGAGCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 418

Qy 112 LeuTyrAlaPheGlyLeuThrAspGluValPheAlaAlaAlaThrAlaLysLeuValArg 131
Db 419 CTCTTCAGCATCGCTGT-----GCCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 469

Qy 132 AsnAsnArg-----ArgTyrSerGluAsnTyrMetIleGlyIle 144
Db 470 GCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 520

Qy 145 Ala-----PheSerSerTyrSerSerTyrValPheGlyThrVal 157
Db 521 CCTGCGCTGTGCTTACCTGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 571

Qy 158 IleGlyAlaPheSerGlySerGlyLeuLeuGlnGlyTyrProAlaValGluAlaAla--- 176
Db 572 -----CTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 604

Qy 177 ---LeuGlyPheMetLeuProAlaLeuPheMetSerPheMetLeuAlaSerPheGlnArg 195

```


Qy 42 ArgLeuGly-----PheSerProLeuGluSerValPhe-----Phe 53
 Db 2115 CGGTTCGGCGCGGTGTTTCGGCTGCTGAACGCTGTTTCGGCAACGGCGGGTTC 2056
 Qy 54 SerCysIleIleIleValAlaGlyAlaSerGlnPheValIleThrAlaMetLeuAlaGly 73
 Db 2055 CGGTGC-----GGTTGGCCACGTTCTGCGCTGCGGTTCGGCGCGGTTC 2008
 Qy 74 SerSerLeuTrpIle-----AlaAlaLeuThr 82
 Db 2007 TCCGGTTTGACCTTCGGTTTCGGCGGTGCGGCATCTGCAACGGTTCGGCGGTTC 1948
 Qy 83 ValMetAlaMetAspValArgHisValLeuThrGlyProSerLeuArgSerArgIleIle 102
 Db 1947 GTTGGCGCTTTGGATTTCGGCTTCGTT----- 1921
 Qy 103 GlnArgLeuGlnLysSerLysThrAlaLeuTrpAlaPheGlyLeuThrAspGluValPhe 122
 Db 1920 -----GGCGCGTTTCGGTTCGGCAGCGTTCGCGCGTTC 1891
 Qy 123 AlaAlaAlaThrAlaLysLeuValArgAsn----- 132
 Db 1890 GTTGGCGCGCGGTTCGGTTTCGGCTTCGGCTTCGGCGCGGTTCGCTTCGTC 1831
 Qy 133 -----AsnArgArgTrpSerGluAsnTrpMetIleGlyIleAlaPheSerSerTrp 149
 Db 1830 GGCAGTTTTCGGCGCGGTTCGGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCT 1777
 Qy 150 SerSerTrpValPheGlyThrValIleGlyAlaPheSer----- 162
 Db 1776 GCGCGCGCGGTTCGGCGCGGTTCGGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCG 1717
 Qy 163 -----GlySerGlyLeuGlnGlyTrpProAlaValGluAlaAlaLeu 177
 Db 1716 TTTTTCGAGGTTTCGGCAGCGGTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCG 1657
 Qy 178 -----GlyPheMetLeu 181
 Db 1656 GAGCATGCTTTGAAGCTGTCACCGAGGAGTTTTCCTC 1615

RESULT 7

US-10-467-657-7353/c
 ; Sequence 7353, Application US/10467657
 ; Publication No. US20050260581A1
 ; GENERAL INFORMATION:
 ; APPLICANT: CHIRON SpA
 ; APPLICANT: FONTANA Maria Rita
 ; APPLICANT: PIZZA Mariagrazia
 ; APPLICANT: MASIGNANI Vega
 ; APPLICANT: MONACI Elisabetta
 ; TITLE OF INVENTION: CONOCOCCAL PROTEINS AND NUCLEIC ACIDS
 ; FILE REFERENCE:
 ; CURRENT APPLICATION NUMBER: US/10/467,657
 ; PRIOR FILING DATE: 2003-08-11
 ; PRIOR APPLICATION NUMBER: GB-0103424.8
 ; PRIOR FILING DATE: 2001-02-12
 ; NUMBER OF SEQ ID NOS: 9218
 ; SOFTWARE: SeqWin99, version 1.04
 ; SEQ ID NO 7353
 ; LENGTH: 2763
 ; TYPE: DNA
 ; ORGANISM: Neisseria gonorrhoeae
 US-10-467-657-7353

Alignment Scores:
 Pred. No.: 29.2 Length: 2763
 Score: 93.00 Matches: 54
 Percent Similarity: 33.2% Conservative: 17
 Best Local Similarity: 25.2% Mismatches: 57
 Query Match: 7.5% Indels: 86
 DB: 8 Gaps: 10

US-10-073-293A-4 (1-245) x US-10-467-657-7353 (1-2763)

Qy 27 IleSerTyrlleProValAlaPheAlaPheGlyLeuAsnAla-----Thr 41
 Db 2190 GTCCGTTTCGGCGGTTCGGTTTCGGATTTCGGCTTCGCTTCGCTTCGCGGTATA 2131
 Qy 42 ArgLeuGly-----PheSerProLeuGluSerValPhe-----Phe 53
 Db 2130 CGGTTCGGCGCGGTTCGGCTTCGCTTCGCAACGGCTTCGGCAACGGCGCGGTTC 2071
 Qy 54 SerCysIleIleIleValAlaGlyAlaSerGlnPheValIleThrAlaMetLeuAlaGly 73
 Db 2070 CGGTGC-----GGTTGGCCACGTTCTGCGCTTCGCTTCGCGCGGTTC 2023
 Qy 74 SerSerLeuTrpIle-----AlaAlaLeuThr 82
 Db 2022 TCCGGTTTGACCTTCGGTTTCGGCGGTGCGGCATCTGCAACGGTTCGGCGGTTC 1963
 Qy 83 ValMetAlaMetAspValArgHisValLeuThrGlyProSerLeuArgSerArgIleIle 102
 Db 1962 GTTGGCGCTTTGGATTTCGGCTTCGTT----- 1936
 Qy 103 GlnArgLeuGlnLysSerLysThrAlaLeuTrpAlaPheGlyLeuThrAspGluValPhe 122
 Db 1935 -----GGCGCGTTTCGGTTCGGCAGCGTTCGCGCGTTC 1906
 Qy 123 AlaAlaAlaThrAlaLysLeuValArgAsn----- 132
 Db 1905 GTTGGCGCGCGGTTCGGTTTCGGCTTCGGCTTCGGCGCGGTTCGCTTCGTC 1846
 Qy 133 -----AsnArgArgTrpSerGluAsnTrpMetIleGlyIleAlaPheSerSerTrp 149
 Db 1845 GGCAGTTTTCGGCGCGGTTCGGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCT 1792
 Qy 150 SerSerTrpValPheGlyThrValIleGlyAlaPheSer----- 162
 Db 1791 GCGCGCGGTTCGGCGCGGTTCGGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCG 1732
 Qy 163 -----GlySerGlyLeuGlnGlyTrpProAlaValGluAlaAlaLeu 177
 Db 1731 TTTTTCGAGGTTTCGGCAGCGGTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTC 1672
 Qy 178 -----GlyPheMetLeu 181
 Db 1671 GAGCATGCTTTGAAGCTGTCACCGAGGAGTTTTCCTC 1630

RESULT 8

US-10-873-528-229
 ; Sequence 229, Application US/10873528
 ; Publication No. US20050276814A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Microbial Technics Limited
 ; APPLICANT: Gilbert, Christophe FG
 ; APPLICANT: Hansbro, Philip M
 ; TITLE OF INVENTION: Proteins
 ; FILE REFERENCE: PWC/P21129WO
 ; CURRENT APPLICATION NUMBER: US/10/873,528
 ; CURRENT FILING DATE: 2004-06-23
 ; PRIOR APPLICATION NUMBER: US/09/769,787
 ; PRIOR FILING DATE: 2001-01-26
 ; PRIOR APPLICATION NUMBER: GB 9816337.1
 ; PRIOR FILING DATE: 1998-03-27
 ; PRIOR APPLICATION NUMBER: US 60/125164
 ; PRIOR FILING DATE: 1999-03-19
 ; NUMBER OF SEQ ID NOS: 388
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 229
 ; LENGTH: 1299
 ; TYPE: DNA
 ; ORGANISM: Streptococcus pneumoniae
 US-10-873-528-229

Alignment Scores:

Pred. No.: 17.2 Length: 1299

Score: 91.50 Matches: 61
Percent Similarity: 37.6% Conservative: 39
Best Local Similarity: 22.9% Mismatches: 83
Query Match: 7.3% Indels: 83
DB: Gaps: 15

US-10-073-293A-4 (1-245) x US-10-873-528-229 (1-1299)

Qy 30 IleProValAlaPheAlaPheGlyLeuAsnAlaThrArgLeuGlyPheSerProLeuGlu 49
::: ||| ||| |||
Db 511 GTTCCACAGCTATTGCCAACACAGTTGAAGCA-----ATGATTCACGACTTT 558
Qy 50 SerValPhePheSerCysIlelleTyAlaGlyAlaSerGlnPheValIleThrAlaMet 69
::: ||| ||| |||
Db 559 GTAATTTCTTATCTTCATGATGTA-----TATATTTGGCGAAGTCA 603
Qy 70 LeuAlaAlaGlySerSer-----LeuTrpIleAlaAlaLeuThrValMetAlaMet 86
||| ||| :::
Db 604 TTGACTAATGC CGGAACATTCATAGAATAATGATTTATTCGTCTATTCAGTTCCGTGGCAA 663
Qy 87 AspValArgHisValLeuTyTrpGlyProSerLeuArgSerArgIleileGlnArgLeuGln 106
::: ||| ||| |||
Db 664 GGTTAACTGGAATCTTGATGATGTCATTGGAAATGCATTCCTTATA----- 711
Qy 107 LysSerLysThrAlaLeuTrpAlaPheGlyLeuThrAspGlu-----ValPheAlaAla 124
::: ||| ||| |||
Db 712 -----TCATTTTTGTGTTGGTTGTTTCATGGCAATCGGTAGTAGTAATGGAGTA 762
Qy 125 AlaThrAlaLysLeuValArgAsn----- 132
|||| | ||| |||
Db 763 GTGACAGCTCTGCTTTTATCTTAATCTTGATGCTAATAAAGCTATGTTAGCCTCTGCTAAT 822
Qy 133 -----AsnArgArgTrpSerGluAsnTrpMetIle 142
::: ||| ||| |||
Db 823 CTATCATAGAAAATGTTGCACATATTTGTTACTCAACAATTTTAGATTCATTTTAAT 882
Qy 143 -----GlyIleAlaPheSerSerTrpSerSerTrpValPheGlyThrValIle 158
::: ||| ||| |||
Db 883 CTATCAGGTTTCAGGGAITACGTTTGGTCTGTAGTGGCCATGCTTTTTCAGCAAAATCA 942
Qy 159 GlyAlaPheSerGlySerGlyLeuLeuGlnGlyTyPrProAlaVal----- 173
::: ||| ||| |||
Db 943 AAACAATPACCAAGCTTTAGGAAAAGTTGCAGCTTTTCCAGCAATATTTAACGTAATGAG 1002
Qy 174 GluAlaAlaLeuGlyPhe-----MetLeuProAlaLeuPheMetSerPheLeuLeu 190
||| ||| ||| |||
Db 1003 CCAGTTGTTATTTGGAATTCGATTCGATGAATCCAGTTAAGTTTGACCTTCACTTCT 1062
Qy 191 AlaserPheGlnArgLysGlnSerLeuCysValThrAlaAlaLeuVal-----GlyAla 208
||| ||| ||| |||
Db 1063 GTTCT-----GTACTTCAGCTGTGATAGTATATGAGCT 1098
Qy 209 LeuAla-----GlyValThrLeu---PheSerIleProVal 219
::: ||| ||| |||
Db 1099 ATTGCAACAGGTTTCATGCAGCCATTCTCAGGGGTAACTTGCCTTGGAGTACACCA-- 1155
Qy 220 AlaIleLeuAlaGlyIleValCysGlyCysLeuThrAlaLeulleGlnAlaPheTrpGln 239
||| ||| ||| |||
Db 1156 GCATTTTATCAGGAATTTTGGTGGGT-----GGATGGCAA 1191
Qy 240 GlyAlaProaspGluLeu 245
||| ||| |||
Db 1192 GGAGTTATTACTCAGCTG 1209

RESULT 9
US-10-793-626-1059
; Sequence 1059, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793.626

```

; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1059
; LENGTH: 1947
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-1059

Alignment Scores:
Pred. No.:      27.6      Length:      1947
Score:          91.50     Matches:      63
Percent Similarity: 32.3%   Conservative: 34
Best Local Similarity: 21.0% Mismatches:    98
Query Match:      7.3%     Indels:      105
DB:              8        Gaps:       15

US-10-073-293A-4 (1-245) x US-10-793-626-1059 (1-1947)

Qy 9 AlaProGlySerAlaThrPheMetGluGlyCysLysAspSerLeuProIleValIleSer 28
Db 466 GCACCACTTGGTGGTAATTCAGTCACAGGT-----CCAGGTGTTAACTAT 510
Qy 29 TyrIleProValAlaPheAlaPheGlyLeuAsnAlaThrArgLeuGlyPheSerProLeu 48
Db 511 TATTTAATGCCAATTCAAATATCTGGTATCGGATCGTTAATGATCGTATCAAC----- 564
Qy 49 GluSerValPhePheSerCysIleIleTyrAla----- 59
Db 565 -----TTCCTTGTACATCTTAAGATGTAAACATCCCAACAATGAAGTTTATGCAA 615
Qy 60 -----GlyAlaSerGlnPheValIleThrAlaMetLeuAlaIaGlySerSer 75
Db 616 ATGCCAATGTTCAAGTGTAAACAACATCATTTACCAACATTAATCTGTATATTAGCATTCCCA 675
Qy 76 LeuTrpIleAlaIalaLeuThrValMetAlaMetAspValArgHisValLeuTyrGlyPro 95
Db 676 GTGTTCACCTGTAGCAGCTGCTTTAATGACTGCTGAT-----AGAAATTTTGGA--- 723
Qy 96 SerLeuArgSerArgIleIleGlnArgLeuGlnLysSerLysThrAlaLeuTrp----- 113
Db 724 -----ACTCAGTCTTCCTCAGTAGCAAAATGGCGGTATGCCAATGCTTTGGGCAAC 774
Qy 114 -----AlaPheGlyLeu 117
Db 775 TTCTTCTGGGTATGGGGGCCACCCCTGAAGTTTATATCGTTATTTTGCAGCATTCGGGTATG 834
Qy 118 ThrAspGluValPheAlaAlaIaThrAlaLysLeuValArgAsnAsnArgArgTrpSer 137
Db 835 TACTCAGAAATCATCCCTACTTTTGGCCGTA---CGTTATTTCGGT 879
Qy 138 GluAsnTrpMetIle-----GlyIleAlaPheSerSerTrpSerSerTrpVal 153
Db 880 CATCAAGATATGATTTGGGCAACATGCAGGATATCGCATCTTAAGTTTCTTAGTTTGGTT 939
Qy 154 Phe-----GlyThrValIleGlyAla---PheSerGlySerGly 165
Db 940 CACCATTCTTCATATGGGTAAATGGTCGTAAATTAACCTCATCTTCTCTATCTCAACA 999
Qy 166 LeuLeuGlnGlyTyrProAla-----ValGlu 174
Db 1000 ATGTTAATCGGTGTTCCAACGGGAGTTAAACTATTTAACTGGTGTCTCATTATACAA 1059
Qy 175 AlaAlaLeuGlyPheMetLeuProAlaLeuPheMetSerPheLeuLeuAlaSerPheGln 194
Db 1060 GGTAGAATATACATTTAGTGCACCTATGCTATTTCTCATTAGCATTCATCCCTAACTTC--- 1116
Qy 195 ArgLysGlnSerLeuCysValThrAlaAlaLeuValGlyAlaLeuAlaGlyVal----- 212

```

```
Db 1117 -----||||:|||||:|||||
; TTTAGGAGGGTTACTGGTGAATGCTT 1146
Qy 213 -----ThrLeuPheSerIlePro----- 218
; |||||:|||||:|||||
; |||||:|||||:|||||
Db 1147 GCAATGGCATCAGCTGACTATCAATATACACACTTATTCTTAGTAGCTCACTTCCAC 1206
; |||||:|||||:|||||
; |||||:|||||:|||||
Qy 219 ValAlaLeuAlaGlyIleValCysGlyCysLeuThrAlaLeuIleGlnAlaPheTrp 238
; |||||:|||||:|||||
; |||||:|||||:|||||
Db 1207 TATACATTGGTTACTGGTGTAGTATTGCTGCTTAGCTGTTTAATC-----TTCTGG 1260
; |||||:|||||:|||||
; |||||:|||||:|||||
RESULT 10
US-10-793-626-3342/c
; Sequence 3342, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: P03480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; PRIOR FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3342
; LENGTH: 2397
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-3342
Alignment Scores:
Pred. No.: 35.2 Length: 2397
Score: 91.50 Matches: 63
Percent Similarity: 32.3% Conservative: 34
Best Local Similarity: 21.0% Mismatches: 98
Query Match: 7.3% Indels: 105
DB: 8 Gaps: 15
US-10-073-293A-4 (1-245) x US-10-793-626-3342 (1-2397)
Qy 9 AlaProGlySerAlaThrPheMetGluGlyCysLysAspSerLeuProIleValIleSer 28
; |||||:|||||:|||||
; |||||:|||||:|||||
; |||||:|||||:|||||
Db 2066 GCACCCTTCTGCTGGTGAATTCAGTCCAGT-----CCAGGTGTTAATCAT 2022
; |||||:|||||:|||||
; |||||:|||||:|||||
; |||||:|||||:|||||
Qy 29 TyrIleProValAlaPheAlaPheGlyLeuAsnAlaThrArgLeuGlyPheSerProLeu 48
; |||||:|||||:|||||
; |||||:|||||:|||||
; |||||:|||||:|||||
Db 2021 TATTAAATTGCAATTCAAATATCTGTTATCGGATCGTTAATGACTGGTATCAAC----- 1968
; |||||:|||||:|||||
; |||||:|||||:|||||
; |||||:|||||:|||||
Qy 49 GluSerValPhePheSerCysIleIleTyrAla----- 59
; |||||:|||||:|||||
; |||||:|||||:|||||
; |||||:|||||:|||||
Db 1967 -----TTCTTTGTTACGATTCTAAGATGTAAACTCCAAATGAAGTTTATGCAA 1917
; |||||:|||||:|||||
; |||||:|||||:|||||
; |||||:|||||:|||||
Qy 60 -----GlyAlaSerGlnPheValIleThrAlaMetLeuAlaAlaGlySerSer 75
; |||||:|||||:|||||
; |||||:|||||:|||||
; |||||:|||||:|||||
Db 1916 ATGCCAATGTCAGTGTACACACATTCATTACACATTAATCGTTATATTAGCAATCCCA 1857
; |||||:|||||:|||||
; |||||:|||||:|||||
; |||||:|||||:|||||
Qy 76 LeuTrpIleAlaAlaLeuThrValMetAlaMetAspValArgHisValLeuTyrGlyPro 95
; |||||:|||||:|||||
; |||||:|||||:|||||
; |||||:|||||:|||||
Db 1856 GTGTTCACTGTAGCATTGCTTTAATGACTGCTGAT-----AGAAATTTTGA--- 1809
; |||||:|||||:|||||
; |||||:|||||:|||||
; |||||:|||||:|||||
Qy 96 SerLeuArgSerArgIleIleGlnArgLeuGlnLysSerLysThrAlaLeuTrp----- 113
; |||||:|||||:|||||
; |||||:|||||:|||||
; |||||:|||||:|||||
Db 1808 -----ACTCAGTTCTTCACTGTAGCAAAATGGCGGTATGCAATGCTTTGGCAAAC 1758
; |||||:|||||:|||||
; |||||:|||||:|||||
; |||||:|||||:|||||
Qy 114 -----AlaPheGlyLeu 117
; |||||:|||||:|||||
; |||||:|||||:|||||
; |||||:|||||:|||||
Db 1757 TTCTTCTGGGTATGGGGCCACCCCTGAAGTTTATATCGTTATTGTCAGCAATTCGGTATG 1698
; |||||:|||||:|||||
; |||||:|||||:|||||
; |||||:|||||:|||||
Qy 118 ThrAspGluValPheAlaAlaAlaThrAlaLysLeuValArgAsnAsnArgArgTrpSer 137
; |||||:|||||:|||||
; |||||:|||||:|||||
; |||||:|||||:|||||
```

```
Db 1697 TACTCAGAAATCATCCCTACTTTTGGCCGTAATA-----CGTTTATCGGT 1653
; |||||:|||||:|||||
; |||||:|||||:|||||
; |||||:|||||:|||||
Qy 138 GluAsnTrpMetIle-----GlyIleAlaPheSerSerTrpSerTrpVal 153
; |||||:|||||:|||||
; |||||:|||||:|||||
; |||||:|||||:|||||
Db 1652 CATCAAGTATGATTGGGCACTGCGAGGATCCATCTTAAGTTTCTTAGTTGGTT 1593
; |||||:|||||:|||||
; |||||:|||||:|||||
; |||||:|||||:|||||
Qy 154 Phe-----GlyThrValIleGlyAla-----PheSerGlySerGly 165
; |||||:|||||:|||||
; |||||:|||||:|||||
; |||||:|||||:|||||
Db 1592 CACCATTCTTCACTATGCGTAAATTAACCTCATCTTCTCTATCTCAACA 1533
; |||||:|||||:|||||
; |||||:|||||:|||||
; |||||:|||||:|||||
Qy 166 LeuLeuGlnGlyTyrProAla-----ValGlu 174
; |||||:|||||:|||||
; |||||:|||||:|||||
; |||||:|||||:|||||
Db 1532 ATGTTAATCGGTGTTCCAAACCGGAGTTAAACTATTAACTGGTGTGCACATTATACAAA 1473
; |||||:|||||:|||||
; |||||:|||||:|||||
; |||||:|||||:|||||
Qy 175 AlaAlaLeuGlyPheMetLeuProAlaLeuPheMetSerPheLeuLeuAlaSerPheGln 194
; |||||:|||||:|||||
; |||||:|||||:|||||
; |||||:|||||:|||||
Db 1472 GGTAGAATTACATTTGAGTCACCTTATCTCATTAGCATTCATCCCTAACTTC--- 1416
; |||||:|||||:|||||
; |||||:|||||:|||||
; |||||:|||||:|||||
Qy 195 ArgLysGlnSerLeuCysValThrAlaAlaLeuValGlyAlaLeuAlaGlyVal--- 212
; |||||:|||||:|||||
; |||||:|||||:|||||
; |||||:|||||:|||||
Db 1415 -----TTATTAGAGGGGTACTGGTGAATGCTT 1386
; |||||:|||||:|||||
; |||||:|||||:|||||
; |||||:|||||:|||||
Qy 213 -----ThrLeuPheSerIlePro----- 218
; |||||:|||||:|||||
; |||||:|||||:|||||
; |||||:|||||:|||||
Db 1385 GCAATGGCATCAGCTGACTATCAATATCAACACACTTATTCTTAGTAGCTCACTCCAC 1326
; |||||:|||||:|||||
; |||||:|||||:|||||
; |||||:|||||:|||||
Qy 219 ValAlaIleLeuAlaGlyIleValCysGlyCysLeuThrAlaLeuIleGlnAlaPheTrp 238
; |||||:|||||:|||||
; |||||:|||||:|||||
; |||||:|||||:|||||
Db 1325 TATACATTGGTTACTGGTGTAGTATTGCTGCTTAGCTGGTTTAATC-----TTCTGG 1272
; |||||:|||||:|||||
; |||||:|||||:|||||
; |||||:|||||:|||||
RESULT 11
US-10-793-626-3792/c
; Sequence 3792, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: P03480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3792
; LENGTH: 3298
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-3792
Alignment Scores:
Pred. No.: 51.1 Length: 3298
Score: 91.50 Matches: 63
Percent Similarity: 32.3% Conservative: 34
Best Local Similarity: 21.0% Mismatches: 98
Query Match: 7.3% Indels: 105
DB: 8 Gaps: 15
US-10-073-293A-4 (1-245) x US-10-793-626-3792 (1-3298)
Qy 9 AlaProGlySerAlaThrPheMetGluGlyCysLysAspSerLeuProIleValIleSer 28
; |||||:|||||:|||||
; |||||:|||||:|||||
; |||||:|||||:|||||
Db 1881 GCACCCTTCTGCTGGTGAATTCAGTCCAGT-----CCAGGTGTTAATCAT 1837
; |||||:|||||:|||||
; |||||:|||||:|||||
; |||||:|||||:|||||
Qy 29 TyrIleProValAlaPheAlaPheGlyLeuAsnAlaThrArgLeuGlyPheSerProLeu 48
; |||||:|||||:|||||
; |||||:|||||:|||||
; |||||:|||||:|||||
Db 1836 TATTAAATTGCAATTCAAATATCTGTTATCGGATCGTTAATGACTGGTATCAAC----- 1783
; |||||:|||||:|||||
; |||||:|||||:|||||
; |||||:|||||:|||||
Qy 49 GluSerValPhePheSerCysIleIleTyrAla----- 59
; |||||:|||||:|||||
; |||||:|||||:|||||
; |||||:|||||:|||||
```


Db 24800 TTTCTTCTGGAAGCTTCATCCC 24821
RESULT 15
US-11-121-086-3
; Sequence 3, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121.086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 3
; LENGTH: 168516
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-3

Alignment Scores:
Pred. No.: 8 08e+03 Length: 168516
Score: 89.50 Matches: 58
Percent Similarity: 35.6% Conservative: 32
Best Local Similarity: 22.9% Mismatches: 104
Query Match: 7.2% Indels: 60
DB: 12 Gaps: 8

US-10-073-293A-4 (1-245) x US-11-121-086-3 (1-168516)

Qy 6 ProGlnProAlaProGlySerAlaThrPheMetGluGlyCysLysAspSerLeuProIle 25
Db 9486 CCGCACCTTCCCTAGCGGCCACCTGCTCCCTCCCGTACGGCTGTAGTCC 9545
Qy 26 ValIleSerTyrIleProVal----- 32
Db 9546 TCTGTGACCCCTGTTCACATCAGCTGACCTGGAAGTTGGCTGTCCCAACAGCAGGC 9605
Qy 33 ---AlaPheAlaPheGlyLeuAsnAlaThrArgLeuGlyPheSerProLeuGluSerVal 51
Db 9606 GCTGCTGGGGCCCAAGGGCTTGTCTCAGGAGGC--GGTTCCCGACCTCTGAT----- 9657
Qy 52 PhePheSerCysIleIleTyrAlaGlyAlaSerGlnPheValIleThrAlaMetLeuAla 71
Db 9658 -----CCCCAGCAACGTCCTCCATGTAACTGGTTCTCTGTA 9696
Qy 72 AlaGlySerSerLeu-----TrrIleAlaAlaLeuThrVal 83
Db 9697 CAGGCACTGGCTTGTGGGCCGCCAGGCTCTCCCATGGCTGGAGGTGCTCACTGTC 9756
Qy 84 MetAlaMetAspValArgHisValLeuTyrGlyProSerLeuArgSerArgIleIleGln 103
Db 9757 TGTGAGGCCCTCGCCAGGCCCTGTGTCCAGAGCCTTCCTGGGCTTGGGGTCTTCCA 9816
Qy 104 ArgLeuGlnLysSerLys-ThrAlaLeuTrrAlaPheGlyLeuThrAspGluValPheAl 123
Db 9817 GGTCTGGCCAGGCCCTCTGTGTGTGGCCATCTGCACCCCATCCCCA-----GC 9870
Qy 123 aAlaAlaThrAlaLysLeuValArgAsnArgTrpSerGluAsnTrpMetIleG1 143
Db 9871 CACATCCTTGGCGCAGCTGTGGAGGATCTGAGGCGCTGGGGTAAATGACTGATGCCAGG 9930
Qy 143 yIleAlaPheSerSerTrrPheSerTrrP-----ValPheGlyTh 156
Db 9931 GTCAGGTTGTCACTGACGGCTCGACGAGGTGAACGCCCCCGCTCTCAGTCTTCCCTTC 9990
Qy 156 rValIleGlyAlaPheSerGlySerGlyLeuGlnGlyTyrProAlaValGluAlaAl 176
Db 9991 TGTCTGTGGCGCTCACGGTGGGCCCTCCCTGAGTGGGTGCGCCACGCTG---GCCGG 10047

Search completed: February 16, 2006, 18:26:38
Job time : 705.857 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 12:15:47 ; Search time 47.0815 Seconds
(without alignments)
1035.886 Million cell updates/sec

Title: US-10-073-293A-6
Perfect score: 552
Sequence: 1 MSTEVLLGLLVGVNYCFR.....IPTLLSALAYGLAWKNVAIL 111

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues
Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : A_Geneseq_21.*
- 1: Geneseqp1980s.*
 - 2: Geneseqp1990s.*
 - 3: Geneseqp2000s.*
 - 4: Geneseqp2001s.*
 - 5: Geneseqp2002s.*
 - 6: Geneseqp2003as.*
 - 7: Geneseqp2003bs.*
 - 8: Geneseqp2004s.*
 - 9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	552	100.0	111	5	ABP57757 E. coli L
2	552	100.0	111	9	ADZ39953 E.coli b2
3	552	100.0	111	9	ADZ36197 Escherich
4	403	73.0	123	7	ABO63701 Klebsiell
5	231	41.8	74	4	RAM84790 Human imm
6	182.5	33.1	122	6	ADA35577 Acinetoba
7	78	14.1	530	6	ABU35859 Protein e
8	77.5	14.0	442	6	ABU35020 Protein e
9	77	13.9	308	8	ADN20876 Bacterial
10	76.5	13.9	494	2	AAZ27289 Glucose t
11	76.5	13.9	500	8	ADS23412 Bacterial
12	76	13.8	463	3	AAZ04739 Arabidops
13	76	13.8	502	3	AAZ04738 Arabidops
14	76	13.8	509	5	ABU65197 Human NOV
15	76	13.8	509	8	ADN62045 Human nov
16	76	13.8	561	3	AAZ04737 Arabidops
17	76	13.8	650	6	ABU42744 Protein e
18	76	13.8	655	5	ABP40686 Staphyloc
19	76	13.8	655	8	ADS07960 Staphyloc
20	75.5	13.7	201	6	ADB11930 Alloiooc
21	75.5	13.7	225	6	ADB11928 Alloiooc
22	75.5	13.7	253	6	ADB11926 Alloiooc
23	75.5	13.7	254	6	ADB11924 Alloiooc
24	75.5	13.7	283	6	ADB11922 Alloiooc

25	75.5	13.7	326	6	ADB11920
26	75.5	13.7	327	6	ADB11918
27	75.5	13.7	497	8	ADQ66449 Novel hum
28	75.5	13.7	497	9	ADX87332 Human glu
29	75.5	13.7	520	8	ADQ66738 Novel hum
30	74	13.4	202	4	AAZ78985 C. glutam
31	74	13.4	202	4	AAZ78984 C. glutam
32	74	13.4	472	5	ABU05729 M. tuberc
33	74	13.4	791	4	AAZ90776 C. glutami
34	74	13.4	791	7	ADL65603 C. glutam
35	73.5	13.3	156	2	AAZ90770 CROC-4 C-
36	73.5	13.3	156	2	AAZ46893 Amino aci
37	73.5	13.3	156	6	ABR58649 Human can
38	73.5	13.3	156	8	ABM80157 Tumour-as
39	73.5	13.3	156	9	ADV70187 Tumour-as
40	73.5	13.3	332	7	ABO62813 Klebsiell
41	73	13.2	269	5	ABZ47774 Listeria
42	73	13.2	298	8	ADV88288 Streptoco
43	73	13.2	298	8	ADV81710 Streptoco
44	73	13.2	298	8	ADV79541 Streptoco
45	73	13.2	319	5	ABP29846 Streptoco

ALIGNMENTS

RESULT 1
ABP57757
ID ABP57757 standard; protein; 111 AA.
XX
AC ABP57757;
XX
DT 29-JAN-2003 (first entry)
XX
DE E. coli L-amino acid producing protein #2.
XX
KW L-amino acid; E. coli.
XX
OS Escherichia coli.
XX
PN EP1239041-A2.
XX
PD 11-SEP-2002.
XX
PF 13-FEB-2002; 2002EP-00003335.
XX
PR 13-FEB-2001; 2001RU-00103865.
PR 26-FEB-2001; 2001RU-00104998.
PR 26-FEB-2001; 2001RU-00104999.
PR 28-JUN-2001; 2001RU-00117632.
PR 28-JUN-2001; 2001RU-00117633.
XX (AJIN) AJINOMOTO CO INC.

Tabolina EA, Rybak KV, Khourges EM, Voroshilova EB, Gussyatiner MM;
WPI; 2002-699856/76.
N-PSDB; ABV75657.
Novel L-amino acid producing Escherichia bacterium, is modified to enhance L-amino acid production by enhancing the activities of protein capable of making bacterium to have enhanced resistance to L-amino acids.
Disclosure; Page 24-25; 33pp; English.
The invention relates to a novel L-amino acid producing bacterium of genus Escherichia, modified to enhance L-amino acid production by enhancing the activities of a protein. The novel bacterium is useful for producing L-amino acid e.g. L-Thr, L-Pro, L-Met and L-Arg, by cultivating the bacterium in a culture medium and collecting L-amino acid to be produced and accumulated from the culture medium. The present sequence represents a protein of the invention which causes increased L-amino acid production in E. coli

XX SQ Sequence 111 AA;

Query Match 100.0%; Score 552; DB 5; Length 111;
 Best Local Similarity 100.0%; Pred. No. 1.5e-60;
 Matches 111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSYEVLGLGLVGVANVCFRYLPLRLRVGNARPTKRGAVGILLDTIGIASICALLVNSTA 60
 Db 1 MSYEVLGLGLVGVANVCFRYLPLRLRVGNARPTKRGAVGILLDTIGIASICALLVNSTA 60

Qy 61 PEVMDTRRFVPTLVGFAVLGASFYKTRSIIPITLLSALAYGLAWKVMII 111
 Db 61 PEVMDTRRFVPTLVGFAVLGASFYKTRSIIPITLLSALAYGLAWKVMII 111

RESULT 2
 ADZ39953
 ID ADZ39953 standard; protein; 111 AA.

XX AC ADZ39953;
 XX DT 14-JUL-2005 (first entry)
 XX DE E.coli b2683 protein for improved amino acid production in bacteria.
 XX KW amino acid production; fermentation; transport protein.
 XX OS Escherichia coli.
 XX PN EP1526179-A1.
 XX PD 27-APR-2005.

XX PF 13-FEB-2002; 2004EP-00028876.
 XX PR 13-FEB-2001; 2001RU-00103865.
 XX PR 26-FEB-2001; 2001RU-00104998.
 XX PR 26-FEB-2001; 2001RU-00104999.
 XX PR 28-JUN-2001; 2001RU-00117632.
 XX PR 28-JUN-2001; 2001RU-00117633.
 XX PR 13-FEB-2002; 2002EP-00003335.
 XX PA (AJIN) AJINOMOTO CO INC.
 XX PI Tabolina EA, Rybak KV, Khourges EM, Voroshilova EB, Gussyatiner MM;
 DR WPI; 2005-308098/32.
 DR N-PSDB; ADZ39952.

XX PT New L-amino acid producing bacterium belonging to the genus Escherichia,
 PT useful for producing L-amino acids, e.g. L-threonine, L-valine, L-
 PT proline, L-methionine, or L-arginine.
 XX Disclosure; SEQ ID NO 6; 35pp; English.

XX CC The invention relates to an L-amino acid producing bacterium belonging to
 CC the genus Escherichia, where the bacterium has been modified so that the
 CC L-amino acid production by the bacterium is enhanced by enhancing
 CC activities of proteins by transformation of the bacterium with DNA coding
 CC for protein or by alteration of an expression regulation sequence of the
 CC DNA on the chromosome of the bacterium. The L-amino acid producing
 CC bacterium is useful for producing L-amino acids by fermentation,
 CC including L-threonine, L-valine, L-proline, L-methionine, or L-arginine.
 CC The genes are useful for improving L-amino acid productivity. This
 CC sequence corresponds to the b2683 protein, one of the L-amino acid
 CC biosynthesis pathway proteins. The protein is a putative transport
 CC protein within the biosynthesis pathway. Enhancement of the expression of
 CC this protein increases production on L-threonine, L-valine, L-leucine, L-
 CC proline and L-methionine.

XX SQ Sequence 111 AA;

Query Match 100.0%; Score 552; DB 9; Length 111;
 Best Local Similarity 100.0%; Pred. No. 1.5e-60;
 Matches 111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSYEVLGLGLVGVANVCFRYLPLRLRVGNARPTKRGAVGILLDTIGIASICALLVNSTA 60
 Db 1 MSYEVLGLGLVGVANVCFRYLPLRLRVGNARPTKRGAVGILLDTIGIASICALLVNSTA 60

Qy 61 PEVMDTRRFVPTLVGFAVLGASFYKTRSIIPITLLSALAYGLAWKVMII 111
 Db 61 PEVMDTRRFVPTLVGFAVLGASFYKTRSIIPITLLSALAYGLAWKVMII 111

RESULT 3
 ADZ36197
 ID ADZ36197 standard; protein; 111 AA.

XX AC ADZ36197;
 XX DT 14-JUL-2005 (first entry)
 XX DE Escherichia coli b2683 protein SEQ ID NO:6.
 XX KW amino acid production.
 XX OS Escherichia coli.
 XX PN EP1526181-A1.
 XX PD 27-APR-2005.

XX PF 13-FEB-2002; 2004EP-00028877.
 XX PR 13-FEB-2001; 2001RU-00103865.
 XX PR 26-FEB-2001; 2001RU-00104998.
 XX PR 26-FEB-2001; 2001RU-00104999.
 XX PR 28-JUN-2001; 2001RU-00117632.
 XX PR 28-JUN-2001; 2001RU-00117633.
 XX PR 13-FEB-2002; 2002EP-00003335.
 XX PA (AJIN) AJINOMOTO CO INC.
 XX PI Tabolina EA, Rybak KV, Khourges EM, Voroshilova EB, Gussyatiner MM;
 DR WPI; 2005-317157/33.
 DR N-PSDB; ADZ36196.

XX PT New modified L-amino acid producing bacterium useful to enhance the
 PT production of L-amino acid by enhancing the activities of proteins in a
 PT cell of bacterium.
 XX Disclosure; SEQ ID NO 6; 35pp; English.

XX CC The invention relates to an L-amino acid (A) producing bacterium (1)
 CC (belonging to the genus Escherichia), which is modified to enhance the
 CC production of (A) by enhancing the activities of proteins (G) or (H) in a
 CC cell of (1). Also described is a method for producing (A) comprising
 CC cultivating the bacterium in a culture medium and collecting the produced
 CC and accumulated L-amino acid. The modified bacterium has the ability to
 CC grow on a minimal medium containing L-amino acid or its analog in a
 CC minimal concentration and ability to grow faster on a medium containing L
 CC -amino acid or its analog than the unmodified strain or the wild type
 CC strain, or the parental strain of the bacterium. The present sequence
 CC represents the E. coli b2683 protein which has L-amino acid excretion
 CC activity.

XX SQ Sequence 111 AA;

Query Match 100.0%; Score 552; DB 9; Length 111;
 Best Local Similarity 100.0%; Pred. No. 1.5e-60;
 Matches 111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSYEVLGLGLVGVANVCFRYLPLRLRVGNARPTKRGAVGILLDTIGIASICALLVNSTA 60

Db 1 MSYEVLLGLLVGVANVCFRYLPRLRVGNARPTKRGAVGILLDTIGIASICALLVVSTA 60
Qy 61 PEVMDTRRRFVPTLVGFPAVLGASFYKTRSIIPITLLSALAYGLAWKVMII 111
Db 61 PEVMDTRRRFVPTLVGFPAVLGASFYKTRSIIPITLLSALAYGLAWKVMII 111

RESULT 4
ABO63701
ID ABO63701 standard; protein; 123 AA.
XX ABO63701;
XX 29-JUL-2004 (first entry)
XX Klebsiella pneumoniae polypeptide seqid 10218.
XX Recombinant expression vector; transcription regulatory element;
KW Klebsiella pneumoniae protein; antibacterial; Vaccine.
XX Klebsiella pneumoniae.
XX US6610836-B1.
XX 26-AUG-2003.
XX 27-JAN-2000; 2000US-00489039.
XX 29-JAN-1999; 99US-0117747P.
XX (GENO-) GENOME THERAPEUTICS CORP.
XX Breton GL, Osborne M;
XX WPI; 2003-895346/82.
XX N-PSDB; ACH97252.
XX New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for
PT preparing a vaccine composition against Klebsiella pneumoniae.
XX Disclosure; SEQ ID NO 10218; 932pp; English.

CC The invention describes a new isolated nucleic acid encoding a Klebsiella
CC pneumoniae polypeptide. Also described are: a recombinant expression
CC vector comprising the nucleic acid, operably linked to a transcription
CC regulatory element; and a cell comprising the recombinant expression
CC vector. The nucleic acid is useful for preparing a vaccine composition
CC against Klebsiella pneumoniae. This is the amino acid sequence of a
CC Klebsiella pneumoniae polypeptide of the invention
XX Sequence 123 AA;
Query Match 73.0%; Score 403; DB 7; Length 123;
Best Local Similarity 72.0%; Pred. No. 6.6e-42;
Matches 77; Conservative 13; Mismatches 17; Indels 0; Gaps 0;
Qy 1 MSYEVLLGLLVGVANVCFRYLPRLRVGNARPTKRGAVGILLDTIGIASICALLVVSTA 60
Db 13 MNSYEVLLGLLVGVANVCFRYLPRLRVGNARPTKRGAVGILLDTIGIASICALLVVSTA 72

Qy 61 PEVMDTRRRFVPTLVGFPAVLGASFYKTRSIIPITLLSALAYGLAWKV 107
Db 73 PEILADAHRLAPTLTGFLVGLGAPFWKTRSIIPITLLSAPAYGLAWKI 119
RESULT 5
AAM84790
ID AAM84790 standard; protein; 74 AA.
XX AAM84790;
XX 07-NOV-2001 (first entry)

XX Human immune/haematopoietic antigen SEQ ID NO:12383.
DE Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytotstatic; gene therapy; vaccine; metastasis.
XX Homo sapiens.
XX WO200157182-A2.
XX 09-AUG-2001.
XX 17-JAN-2001; 2001WO-US001354.
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226988P.
PR 23-AUG-2000; 2000US-0227182P.
PR 30-AUG-2000; 2000US-0227009P.
PR 01-SEP-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 05-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 06-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 08-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231422P.
PR 08-SEP-2000; 2000US-0231443P.
PR 08-SEP-2000; 2000US-0231444P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.

```
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236399P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0246177P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249266P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.

PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-483426/52.
DR N-PSDB; AAK57571.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Claim 11; SEQ ID NO 12383; 307lpp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
XX amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patients own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting the
XX nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/hematopoietic-related diseases, especially
XX cancers and cancer metastases of haematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/hematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAM82169
XX represent sequences used in the exemplification of the present invention
XX
SQ Sequence 74 AA;
Query Match 41.8%; Score 231; DB 4; Length 74;
Best Local Similarity 94.0%; Pred. No. 1e-20;
Matches 47; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 50 SICALLVSTAPEVMDTRFVPTLVGFAVLGASFYKTRSIITPLLSAL 99
DB 20 SICALLVSTAPEVMDTRFVPTLVGFAVLGASFYKTRSIITPLLSAL 69
RESULT 6
ADA35577
ID ADA35577 standard; protein; 122 AA.
XX
XX ADA35577;
AC ADA35577;
DT 20-NOV-2003 (first entry)
XX
DE Acinetobacter baumannii protein #2738.
XX
XX Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;
XX plant biocontrol agent.
XX
XX Acinetobacter baumannii.
XX
XX US6562958-B1.
XX
XX 13-MAY-2003.
XX
XX 04-JUN-1999; 99US-00328352.
XX
XX 09-JUN-1998; 98US-0088701P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Breton G, Bush D;
XX
```


Qy 64 MHDTRRFPVTLVGFVAVLGASFYKTRSTIIIPFLLSALAY 101
 Db 149 AH-----APALIPRTGL-ANFYRGFS----PTLLGMLPY 177

RESULT 10
 ID AAY27289
 AC AAY27289; standard; protein; 494 AA.

DT 05-NOV-1999 (first entry)
 XX Glucose transporter protein GLUT3.
 XX Human; glucose transporter; GLUTX; hexose transport; gene therapy;
 KW chromosome mapping; GLUT3.
 KW Homo sapiens.

XX US5942398-A.
 XX 24-AUG-1999.
 XX 26-FEB-1998; 98US-00031392.
 XX 26-FEB-1998; 98US-00031392.
 XX (MILL-) MILLENNIUM PHARM INC.
 XX Tartaglia LA, Weng X;
 XX WPI; 1999-526192/44.

XX A novel mammalian transporter protein and the gene encoding it, are
 PT useful in the diagnosis and treatment of disorders associated with
 PT aberrant sugar transport.
 XX Example; Fig 3; 48pp; English.

XX The invention relates to a human glucose transporter (GLUTX) protein. The
 CC protein can be expressed by standard recombinant methodology. GLUTX
 CC nucleic acids are useful as hybridization probes for detecting the
 CC presence of GLUTX DNA in a sample, useful for diagnosing conditions
 CC associated with aberrant expression levels of GLUTX. The GLUTX gene is
 CC also useful as a therapeutic agent for regulating translation of GLUTX
 CC mRNA, and for treatment of disorders associated with aberrant expression
 CC of GLUTX and aberrant hexose transport. It is useful for generating GLUTX
 CC specific antibodies, identifying agonists and antagonists of GLUTX, and
 CC identifying nucleic acids in other species encoding nucleic acids
 CC homologous to GLUTX. GLUTX nucleic acids are also useful for identifying
 CC the chromosomal location of GLUTX, and as tissue specific markers.
 CC Sequences AAY27287-291 represent different glucose transporter proteins
 CC which were compared with human GLUTX

XX Sequence 494 AA;
 Query Match 13.9%; Score 76.5; DB 2; Length 494;
 Best Local Similarity 34.1%; Pred. No. 2;
 Matches 29; Conservative 17; Mismatches 30; Indels 9; Gaps 7;

Qy 2 SYEVLLG-LLVGV-ANVCYRLRLRVGNARPTK-RGAVGILLDTIGTIGTICALVWS 58
 Db 116 SVEMLILGRLLIGLFCGLCTGFVP--WYGEISFTALRGAFG-TLNQGLI--VIGILVAQ 170
 Qy 59 T-APEVNMHDTRRFPVTLVGFVAVLGA 82
 Db 171 IFGLKVILGTEDLWPLLGLFTILPA 195

RESULT 11
 ADS23412
 ID ADS23412 standard; protein; 500 AA.

XX ADS23412;
 XX 02-DEC-2004 (first entry)
 XX Bacterial polypeptide #12445.
 XX Recombinant DNA construct; transformed plant; improved plant property;
 KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
 KW pathogen tolerance; pest tolerance; plant disease resistance;
 KW cell cycle pathway modification; plant growth regulator;
 KW homologous recombination; seed oil yield; protein yield; carbohydrate;
 KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
 KW bacterial polypeptide.
 XX Bacteria.
 XX US2003233675-A1.
 XX 18-DEC-2003.
 XX 20-FEB-2003; 2003US-00369493.
 XX 21-FEB-2002; 2002US-0360039P.
 XX (CAOY/) CAO Y.
 XX (HINK/) HINKLE G J.
 XX (SLAT/) SLATER S C.
 XX (CHEN/) CHEN X.
 XX (GOLD/) GOLDMAN B S.
 XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
 XX WPI; 2004-061375/06.
 XX New recombinant DNA construct comprising a promoter positioned to provide
 PT for expression of a polynucleotide encoding a polypeptide from a
 PT microbial source, useful for producing plants with improved properties.
 XX Claim 1; SEQ ID NO 12445; 122pp; English.

XX The invention relates to a recombinant DNA construct comprising a
 CC promoter functional in a plant cell, where the promoter is positioned to
 CC provide for expression of a polynucleotide encoding a polypeptide from a
 CC microbial source. The invention also relates to a transformed plant
 CC comprising the recombinant DNA construct and a method of producing a
 CC transformed plant having an improved property. The plant is a crop plant
 CC such as maize or soybean. The method of producing a transformed plant
 CC having an improved property comprises transforming a plant with the
 CC recombinant DNA construct and growing the transformed plant, where the
 CC polynucleotide or polypeptide is useful for improving plant properties.
 CC The recombinant DNA construct is useful for producing plants with
 CC improved plant properties, e.g. improved cold, heat or drought tolerance,
 CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
 CC increased resistance to plant disease, better growth rate by modification
 CC of the cell cycle pathway with plant growth regulators, increased rate of
 CC homologous recombination, modified seed oil or protein yield and/or
 CC content, improved yield by modification of carbohydrate, nitrogen or
 CC phosphorus use and/or uptake, by modification of photosynthesis or by
 CC providing improved plant growth and development under at least one stress
 CC condition, improved lignin production or improved galactomannan
 CC production. This sequence represents a bacterial polypeptide used in the
 CC scope of the invention. Note: The sequence data for this patent did not
 CC form part of the printed specification but was obtained in electronic
 CC format from USPTO at seqdata.uspto.gov/sequence.html.

XX Sequence 500 AA;
 Query Match 13.9%; Score 76.5; DB 8; Length 500;
 Best Local Similarity 25.2%; Pred. No. 2.1;
 Matches 33; Conservative 20; Mismatches 45; Indels 33; Gaps 5;
 Qy 9 GLLVGVANVCYRYLRLRVGNARPTKRGAVGILLDTIGTIGTICALVWSAPEVMDTR 68

```
Db      20 GLFVGIGAYLRDAGFLSLLEY---LWGIAFILPINLUSVCEMCAYLPIRGS--IFELAA 74
Qy      69 RFVPTLVGFAVLGASFYK---TRS-----IIPTLLSALAYGL----- 103
Db      75 RYVDPAFGFMKXGSAFRVQGRSDNIRLGDWVYFYAGLMLVCTEYSAVAFIMDYWQIDV 134
Qy      104 ---AWKUNAI 111
Db      135 NPAAWVANAMV 145

RESULT 12
AAG04739
ID  AAG04739 standard; protein; 463 AA.
XX
AC  AAG04739;
XX
DT  17-OCT-2000 (first entry)
XX
DE  Arabidopsis thaliana protein fragment SEQ ID NO: 879.
XX
KW  Protein identification; signal transduction pathway; metabolic pathway;
KW  hybridisation assay; genetic mapping; gene expression control; promoter;
KW  termination sequence.
XX
OS  Arabidopsis thaliana.
XX
FN  EP1033405-A2.
XX
PD  06-SEP-2000.
XX
PF  25-FEB-2000; 2000EP-00301439.
XX
PR  25-FEB-1999; 99US-0121825P.
PR  05-MAR-1999; 99US-0123180P.
PR  09-MAR-1999; 99US-0123548P.
PR  23-MAR-1999; 99US-0125788P.
PR  25-MAR-1999; 99US-0126264P.
PR  29-MAR-1999; 99US-0126785P.
PR  01-APR-1999; 99US-0127462P.
PR  06-APR-1999; 99US-0128234P.
PR  08-APR-1999; 99US-0128714P.
PR  16-APR-1999; 99US-0129845P.
PR  19-APR-1999; 99US-0130077P.
PR  21-APR-1999; 99US-0130449P.
PR  23-APR-1999; 99US-0130510P.
PR  28-APR-1999; 99US-0130891P.
PR  30-APR-1999; 99US-0131449P.
PR  30-APR-1999; 99US-0132048P.
PR  30-APR-1999; 99US-0132407P.
PR  04-MAY-1999; 99US-0132484P.
PR  05-MAY-1999; 99US-0132485P.
PR  06-MAY-1999; 99US-0132486P.
PR  06-MAY-1999; 99US-0132487P.
PR  07-MAY-1999; 99US-0132863P.
PR  11-MAY-1999; 99US-0134256P.
PR  14-MAY-1999; 99US-0134218P.
PR  14-MAY-1999; 99US-0134219P.
PR  14-MAY-1999; 99US-0134221P.
PR  14-MAY-1999; 99US-0134370P.
PR  18-MAY-1999; 99US-0134768P.
PR  19-MAY-1999; 99US-0134941P.
PR  20-MAY-1999; 99US-0135124P.
PR  21-MAY-1999; 99US-0135353P.
PR  24-MAY-1999; 99US-0135629P.
PR  25-MAY-1999; 99US-0136021P.
PR  27-MAY-1999; 99US-0136382P.
PR  28-MAY-1999; 99US-0136782P.
PR  01-JUN-1999; 99US-0137222P.
PR  03-JUN-1999; 99US-0137528P.
PR  04-JUN-1999; 99US-0137502P.
PR  07-JUN-1999; 99US-0137724P.
PR  08-JUN-1999; 99US-0138094P.
PR  10-JUN-1999; 99US-0138540P.
PR  10-JUN-1999; 99US-0138847P.
PR  14-JUN-1999; 99US-0139119P.
PR  16-JUN-1999; 99US-0139452P.
PR  16-JUN-1999; 99US-0139453P.
PR  17-JUN-1999; 99US-0139492P.
PR  18-JUN-1999; 99US-0139454P.
PR  18-JUN-1999; 99US-0139455P.
PR  18-JUN-1999; 99US-0139456P.
PR  18-JUN-1999; 99US-0139457P.
PR  18-JUN-1999; 99US-0139458P.
PR  18-JUN-1999; 99US-0139459P.
PR  18-JUN-1999; 99US-0139460P.
PR  18-JUN-1999; 99US-0139461P.
PR  18-JUN-1999; 99US-0139462P.
PR  18-JUN-1999; 99US-0139463P.
PR  18-JUN-1999; 99US-0139750P.
PR  18-JUN-1999; 99US-0139763P.
PR  21-JUN-1999; 99US-0139817P.
PR  22-JUN-1999; 99US-0139899P.
PR  23-JUN-1999; 99US-0140353P.
PR  23-JUN-1999; 99US-0140354P.
PR  24-JUN-1999; 99US-0140695P.
PR  28-JUN-1999; 99US-0140823P.
PR  29-JUN-1999; 99US-0140991P.
PR  30-JUN-1999; 99US-0141287P.
PR  01-JUL-1999; 99US-0141842P.
PR  02-JUL-1999; 99US-0142154P.
PR  02-JUL-1999; 99US-0142055P.
PR  06-JUL-1999; 99US-0142390P.
PR  08-JUL-1999; 99US-0142803P.
PR  09-JUL-1999; 99US-0142920P.
PR  12-JUL-1999; 99US-0142977P.
PR  13-JUL-1999; 99US-0143542P.
PR  14-JUL-1999; 99US-0143624P.
PR  15-JUL-1999; 99US-0144005P.
PR  16-JUL-1999; 99US-0144085P.
PR  16-JUL-1999; 99US-0144086P.
PR  19-JUL-1999; 99US-0144325P.
PR  19-JUL-1999; 99US-0144331P.
PR  19-JUL-1999; 99US-0144332P.
PR  19-JUL-1999; 99US-0144333P.
PR  19-JUL-1999; 99US-0144334P.
PR  19-JUL-1999; 99US-0144335P.
PR  20-JUL-1999; 99US-0144352P.
PR  20-JUL-1999; 99US-0144352P.
PR  20-JUL-1999; 99US-0144632P.
PR  21-JUL-1999; 99US-0144884P.
PR  21-JUL-1999; 99US-0144814P.
PR  21-JUL-1999; 99US-0145086P.
PR  21-JUL-1999; 99US-0145088P.
PR  21-JUL-1999; 99US-0145085P.
PR  22-JUL-1999; 99US-0145087P.
PR  22-JUL-1999; 99US-0145089P.
PR  22-JUL-1999; 99US-0145192P.
PR  23-JUL-1999; 99US-0145145P.
PR  23-JUL-1999; 99US-0145216P.
PR  23-JUL-1999; 99US-0145224P.
PR  26-JUL-1999; 99US-0145276P.
PR  27-JUL-1999; 99US-0145913P.
PR  27-JUL-1999; 99US-0145918P.
PR  27-JUL-1999; 99US-0145919P.
PR  28-JUL-1999; 99US-0145951P.
PR  02-AUG-1999; 99US-0146386P.
PR  02-AUG-1999; 99US-0146388P.
PR  02-AUG-1999; 99US-0146389P.
PR  03-AUG-1999; 99US-0147038P.
PR  04-AUG-1999; 99US-0147204P.
PR  04-AUG-1999; 99US-0147302P.
PR  05-AUG-1999; 99US-0147192P.
PR  05-AUG-1999; 99US-0147260P.
PR  06-AUG-1999; 99US-0147303P.
PR  06-AUG-1999; 99US-0147416P.
```

PR	09-AUG-1999;	99US-0147493P.	QY	6	LLGLLVGVANYCFRYLPLRLVGNAPTRKRGAVGILLD-----TIGIASICAL- 54
PR	09-AUG-1999;	99US-0147935P.	Db	4	LMWGLVGVPTY---YL-----AGSLVDLGMWQGIATVVTANLLLV 44
PR	10-AUG-1999;	99US-0148171P.			
PR	11-AUG-1999;	99US-0148319P.			
PR	12-AUG-1999;	99US-0148341P.			
PR	13-AUG-1999;	99US-0148565P.			
PR	13-AUG-1999;	99US-0148684P.			
PR	16-AUG-1999;	99US-0149368P.			
PR	17-AUG-1999;	99US-0149175P.	QY	55	LVVSTAPEVNHDRFVPTLVGFVAVLGASFYKTRSIILITLSALAYGLAW 105
PR	18-AUG-1999;	99US-0149426P.	Db	45	PLVLTAPQPGTLYG-----ISFPVLARSFGIRGAHIPTLLRALV-GCGW 87
PR	20-AUG-1999;	99US-0149722P.			
PR	20-AUG-1999;	99US-0149723P.			
PR	20-AUG-1999;	99US-0149929P.			
PR	23-AUG-1999;	99US-0149902P.			
PR	23-AUG-1999;	99US-0149930P.			
PR	26-AUG-1999;	99US-0150566P.			
PR	26-AUG-1999;	99US-0150884P.			
PR	27-AUG-1999;	99US-0151085P.			
PR	27-AUG-1999;	99US-0151066P.			
PR	27-AUG-1999;	99US-0151080P.			
PR	30-AUG-1999;	99US-0151303P.			
PR	31-AUG-1999;	99US-0151438P.			
PR	01-SEP-1999;	99US-0151930P.			
PR	07-SEP-1999;	99US-0152363P.			
PR	10-SEP-1999;	99US-0153070P.			
PR	13-SEP-1999;	99US-0153758P.			
PR	15-SEP-1999;	99US-0154018P.			
PR	16-SEP-1999;	99US-0154039P.			
PR	20-SEP-1999;	99US-0154779P.			
PR	22-SEP-1999;	99US-0155139P.			
PR	23-SEP-1999;	99US-0155486P.			
PR	24-SEP-1999;	99US-0155659P.			
PR	28-SEP-1999;	99US-0156458P.			
PR	29-SEP-1999;	99US-0156596P.			
PR	04-OCT-1999;	99US-0157117P.			
PR	05-OCT-1999;	99US-0157753P.			
PR	06-OCT-1999;	99US-0157865P.			
PR	07-OCT-1999;	99US-0158029P.			
PR	08-OCT-1999;	99US-0158232P.			
PR	12-OCT-1999;	99US-0158369P.			
PR	13-OCT-1999;	99US-0159293P.			
PR	13-OCT-1999;	99US-0159294P.			
PR	14-OCT-1999;	99US-0159295P.			
PR	14-OCT-1999;	99US-0159330P.			
PR	14-OCT-1999;	99US-0159331P.			
PR	14-OCT-1999;	99US-0159637P.			
PR	14-OCT-1999;	99US-0159638P.			
PR	18-OCT-1999;	99US-0159584P.			
PR	21-OCT-1999;	99US-0160741P.			
PR	21-OCT-1999;	99US-0160767P.			
PR	21-OCT-1999;	99US-0160768P.			
PR	21-OCT-1999;	99US-0160770P.			
PR	21-OCT-1999;	99US-0160814P.			
PR	21-OCT-1999;	99US-0160815P.			
PR	22-OCT-1999;	99US-0160980P.			
PR	22-OCT-1999;	99US-0160981P.			
PR	22-OCT-1999;	99US-0160989P.			
PR	25-OCT-1999;	99US-0161404P.			
PR	25-OCT-1999;	99US-0161405P.			
PR	25-OCT-1999;	99US-0161406P.			
PR	26-OCT-1999;	99US-0161359P.			
PR	26-OCT-1999;	99US-0161360P.			
PR	26-OCT-1999;	99US-0161361P.			
PR	28-OCT-1999;	99US-0161920P.			
PR	28-OCT-1999;	99US-0161922P.			
PR	28-OCT-1999;	99US-0161931P.			
PR	29-OCT-1999;	99US-0162142P.			

Query Match

13.8%; Score 76; DB 3; Length 463;

Best Local Similarity 28.6%; Pred. No. 2.2;

Matches 32; Conservative 12; Mismatches 28; Indels 40; Gaps 6;

RESULT 13

AAG04738

ID AAG04738 standard; protein; 502 AA.

XX

AC AAG04738;

XX

DT 17-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 878.

XX

KW Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.

KW

XX Arabidopsis thaliana.

OS

XX EP1033405-A2.

PN

XX 06-SEP-2000.

PD

XX 25-FEB-2000; 2000EP-00301439.

XX

XX 25-FEB-1999; 99US-0121825P.

PR

05-MAR-1999; 99US-0123180P.

PR

03-MAR-1999; 99US-0123548P.

PR

23-MAR-1999; 99US-0125788P.

PR

25-MAR-1999; 99US-0126264P.

PR

01-APR-1999; 99US-0127462P.

PR

06-APR-1999; 99US-0128234P.

PR

08-APR-1999; 99US-0128714P.

PR

16-APR-1999; 99US-0129845P.

PR

19-APR-1999; 99US-0130077P.

PR

21-APR-1999; 99US-0130449P.

PR

23-APR-1999; 99US-0130891P.

PR

28-APR-1999; 99US-0131449P.

PR

30-APR-1999; 99US-0132048P.

PR

30-APR-1999; 99US-0132407P.

PR

04-MAY-1999; 99US-0132484P.

PR

05-MAY-1999; 99US-0132485P.

PR

06-MAY-1999; 99US-0132486P.

PR

06-MAY-1999; 99US-0132487P.

PR

07-MAY-1999; 99US-0132863P.

PR

11-MAY-1999; 99US-0134256P.

PR

14-MAY-1999; 99US-0134218P.

PR

14-MAY-1999; 99US-0134219P.

PR

14-MAY-1999; 99US-0134221P.

PR

14-MAY-1999; 99US-0134370P.

PR

18-MAY-1999; 99US-0134768P.

PR

19-MAY-1999; 99US-0134941P.

PR

20-MAY-1999; 99US-0135124P.

PR

21-MAY-1999; 99US-0135353P.

PR

24-MAY-1999; 99US-0135629P.

PR

25-MAY-1999; 99US-0136021P.

PR

27-MAY-1999; 99US-0136392P.

PR

28-MAY-1999; 99US-0136782P.

PR

01-JUN-1999; 99US-0137222P.

PR

03-JUN-1999; 99US-0137528P.

PR

04-JUN-1999; 99US-0137502P.

PR

07-JUN-1999; 99US-0137724P.

PR

08-JUN-1999; 99US-0138094P.

PR

10-JUN-1999; 99US-0138540P.

PR

10-JUN-1999; 99US-0138847P.

PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 17-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139452P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142300P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145114P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.

PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 22-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 13.8%; Score 76; DB 3; Length 502;

Best Local Similarity 28.6%; Pred. No. 2.4; Mismatches 12; Conservative 32; Indels 40; Gaps 6;

Oy 6 LLLGLLVGVANYCFRFLRLRVGNARPTKRGAVGILLD-----TTGIAISICAL- 54

Db 43 LWIGLVGVPTY---YL-----AGSLVDLGMWVQGIATVVTANLILIV 83

KW Human; NOVX; diabetes; obesity; infectious disease; anorexia;
KW cancer-associated cachexia; cancer; neurodegenerative disorder;
KW Alzheimer's disease; Parkinson's disease; immune disorder;
KW haematopoietic disorder; dyslipidaemia; chronic disease.

XX Homo sapiens.

XX US2004043382-A1.

XX 04-MAR-2004.

XX 07-MAR-2002; 2002US-00092900.

XX 08-MAR-2001; 2001US-0274191P.

XX 08-MAR-2001; 2001US-0274194P.

XX 08-MAR-2001; 2001US-0274281P.

XX 09-MAR-2001; 2001US-0274322P.

XX 12-MAR-2001; 2001US-0274849P.

XX 13-MAR-2001; 2001US-0275235P.

XX 13-MAR-2001; 2001US-0275578P.

XX 13-MAR-2001; 2001US-0275579P.

XX 13-MAR-2001; 2001US-0275601P.

XX 14-MAR-2001; 2001US-0276000P.

XX 16-MAR-2001; 2001US-0276776P.

XX 19-MAR-2001; 2001US-0276994P.

XX 20-MAR-2001; 2001US-0277239P.

XX 20-MAR-2001; 2001US-0277321P.

XX 20-MAR-2001; 2001US-0277327P.

XX 20-MAR-2001; 2001US-0277338P.

XX 21-MAR-2001; 2001US-0277791P.

XX 22-MAR-2001; 2001US-0277833P.

XX 23-MAR-2001; 2001US-0278152P.

XX 26-MAR-2001; 2001US-0278894P.

XX 27-MAR-2001; 2001US-0278999P.

XX 27-MAR-2001; 2001US-0279036P.

XX 28-MAR-2001; 2001US-0279344P.

XX 30-MAR-2001; 2001US-0279995P.

XX 30-MAR-2001; 2001US-0280233P.

XX 02-APR-2001; 2001US-0280802P.

XX 02-APR-2001; 2001US-0280802P.

XX 02-APR-2001; 2001US-0280802P.

XX 04-APR-2001; 2001US-0280900P.

XX 13-APR-2001; 2001US-0281444P.

XX 13-APR-2001; 2001US-0281675P.

XX 30-APR-2001; 2001US-0287424P.

XX 02-MAY-2001; 2001US-0288066P.

XX 03-MAY-2001; 2001US-0288342P.

XX 15-MAY-2001; 2001US-0291190P.

XX 16-MAY-2001; 2001US-0291099P.

XX 16-MAY-2001; 2001US-0291240P.

XX 30-MAY-2001; 2001US-0294485P.

XX 31-MAY-2001; 2001US-0294889P.

XX 31-MAY-2001; 2001US-0294899P.

XX 18-JUN-2001; 2001US-0299027P.

XX 19-JUN-2001; 2001US-0299303P.

XX 19-JUN-2001; 2001US-0299310P.

XX 10-JUL-2001; 2001US-0304354P.

XX 31-JUL-2001; 2001US-0309198P.

XX 16-AUG-2001; 2001US-0312903P.

XX 10-SEP-2001; 2001US-0318462P.

XX 12-SEP-2001; 2001US-0318770P.

XX 27-SEP-2001; 2001US-0325430P.

XX 27-SEP-2001; 2001US-0325681P.

XX 18-OCT-2001; 2001US-0330380P.

XX 31-OCT-2001; 2001US-0335301P.

XX 14-NOV-2001; 2001US-0332172P.

XX 14-NOV-2001; 2001US-0332271P.

XX 14-NOV-2001; 2001US-0332272P.

XX 14-NOV-2001; 2001US-0333184P.

XX 14-NOV-2001; 2001US-0333272P.

XX 21-NOV-2001; 2001US-0332094P.

XX 03-DEC-2001; 2001US-0337426P.

XX 03-DEC-2001; 2001US-0338092P.

XX 04-DEC-2001; 2001US-0337185P.

XX 03-JAN-2002; 2002US-0345705P.

XX (PADI/) PADIGARU M.

XX (SPYT) SPYTEK K A.

XX (SHEN) SHENOY S G.

XX (TAUP/) TAUPIER R J.

XX (PENA/) PENNA C E A.

XX (LILL/) LI L.

XX (ZERH/) ZERHUSEN B D.

XX (GUSE/) GUSEV V Y.

XX (JIWW/) JI W.

XX (GORM/) GORMAN L.

XX (MILL/) MILLER C E.

XX (KEKU/) KEKUDA R.

XX (PATT/) PATTURAJAN M.

XX (GANG/) GANGOLLI E A.

XX (VERN/) VERNET C A M.

XX (GUOX/) GUO X S.

XX (TCHE/) TCHERNEV V T.

XX (FERN/) FERNANDES E R.

XX (CASM/) CASMAN S J.

XX (MALY/) MALYANKAR U M.

XX (GERL/) GERLACH V.

XX (LIUY/) LIU Y.

XX (ANDE/) ANDERSON D W.

XX (SPAD/) SPADERNA S K.

XX (CATT/) CATTERTON E.

XX (LEIT/) LEITE M W.

XX (ZHON/) ZHONG H.

XX (ALSO/) ALSOBROOK J P.

XX (LEPL/) LEPLEY D M.

XX (RIEG/) RIEGER D K.

XX (BURG/) BURGESS C E.

XX Padigaru M, Spytek KA, Shenoy SG, Taupier RJ, Pena CEA, Li L; Zerkhusen BD, Gusev VV, Ji W, Gorman L, Miller CE, Kekuda R; Patturajan M, Gangolli EA, Vernet CAM, Guo XS, Tchernev VT; Fernandes ER, Casman SJ, Malyankar UM, Gerlach V, Liu Y; Anderson DM, Spaderna SK, Catterton E, Leite MW, Zhong H; Alsobrook JP, Lepley DM, Rieger DK, Burgess CE; WPI: 2004-225693/21.

XX N-PSDB; ADN62044.

XX New NOVX polypeptides and nucleic acid molecules useful for diagnosing, preventing or treating NOVX-associated disorders, e.g. cancer, diabetes, infection or obesity, and in chromosome mapping, tissue typing or pharmacogenomics.

XX Claim 1; SEQ ID NO 314; 786pp; English.

XX The invention relates to an isolated polypeptide (designated NOVX, or NOV1-NOV127) comprising a sequence selected from 178 fully defined amino acid sequences (and their mature forms, variants and fragments). Also included are an isolated nucleic acid molecule encoding NOVX, a vector comprising the nucleic acid, a cell comprising the vector, methods for determining the presence or amount of the polypeptide or the nucleic acid molecule in a sample, methods for determining the presence of or predisposition to a disease associated with altered levels of expression of the above polypeptide or nucleic acid molecule in a first mammalian subject, a method for identifying an agent that binds to the above polypeptide, a method for identifying a potential therapeutic agent for use in the treatment of a pathology that is related to aberrant expression or physiological interactions of the polypeptide, a method of screening for a modulator of activity or of latency or predisposition to a pathology associated with the polypeptide and a method for modulating the activity of the polypeptide cited above. The composition and methods are useful for diagnosing, preventing or treating diseases such as diabetes, obesity, infectious diseases, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders like Alzheimer's disease or Parkinson's disease, immune disorders, haematopoietic disorders, dyslipidaemias, and other chronic diseases. These may also be used in

CC chromosome mapping, tissue typing, preventive medicine and
CC pharmacogenomics. The polypeptides are also useful as vaccines. The
CC present sequence represents a NOVX protein of the invention.
XX
SQ Sequence 509 AA;
Query Match 13.8%; Score 76; DB 8; Length 509;
Best Local Similarity 32.1%; Pred. No. 2.4;
Matches 27; Conservative 17; Mismatches 34; Indels 6; Gaps 5;
Qy 2 SYEVLLLG-LLVGV-ANVCFRYLPRLRLRVGNARPTK-RGAVGILLDTIGIASICALLVVS 58
Db 119 SVEMLILGRLVIGLFCGLCTGFP--WYGEISPTALRGAPG-TINQLGIVIGILVAQVI 175
Qy 59 TAPEVMHDTRRFVPTLVGFAVLGA 82
Db 176 FGLEILGSBELWPVLLGFTILPA 199
Search completed: February 15, 2006, 12:18:34
Job time : 50.0815 secs

This Page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 12:18:47 ; Search time 14.9663 Seconds
(without alignments)
713.608 Million cell updates/sec

Title: US-10-073-293A-6
Perfect score: 552
Sequence: 1 MSYEVLGLLVGVNYCFR.....IPTLLSALAYGLAWKVMII 111

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_80.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	552	100.0	111	2 D65048	hypothetical prote
2	548	99.3	111	2 A91072	hypothetical prote
3	544	98.6	111	2 D85916	hypothetical prote
4	308.5	55.9	113	2 A90396	probable membrane
5	78.5	14.2	866	2 AG5272	hypothetical prote
6	78	14.1	530	2 G87022	hypothetical prote
7	76.5	13.9	309	2 B69566	phosphate ABC tran
8	76	13.8	599	2 T48383	uracil transporter
9	74.5	13.5	277	1 H71220	hypothetical prote
10	74.5	13.5	486	2 H84805	hypothetical prote
11	74	13.4	472	2 H70836	hypothetical prote
12	73.5	13.3	373	2 A49806	prv43 protein - su
13	73	13.2	269	2 AG1126	conserved membrane
14	73	13.2	269	2 AD1487	light harvesting c
15	73	13.2	459	2 S28025	hydrogenase 4 memb
16	73	13.2	672	2 D85891	glucose transporte
17	72.5	13.1	496	2 A31986	tag protein - Esc
18	72.5	13.1	637	2 T08530	trag protein - Esc
19	72.5	13.1	637	2 S22992	hypothetical prote
20	72	13.0	481	2 B86285	cytochrome c-type
21	72	13.0	660	1 S54746	probable metabolit
22	71.5	13.0	431	2 B96006	probable efflux pu
23	71.5	13.0	1037	2 AE0816	Na+/H+ antiporter
24	71	12.9	388	2 AB0507	conserved hypotet
25	71	12.9	464	2 C75474	hypothetical prote
26	70.5	12.8	440	2 T50912	MFS permease limpo
27	70.5	12.8	542	2 AF2587	hypothetical prote
28	70.5	12.8	542	2 F97369	conserved hypotet
29	70	12.7	100	2 A69469	

30	70	12.7	348	2 B84299	hypothetical prote
31	70	12.7	652	2 B89841	fructose specific
32	70	12.7	672	2 A65024	Hydrogenase-4 comp
33	70	12.7	672	2 H91046	hydrogenase 4 memb
34	69.5	12.6	448	2 B82991	glycerol-3-phospha
35	69.5	12.6	475	2 B29606	methylphenomycin A r
36	69.5	12.6	484	2 F96515	Fl6N3.20 [imported
37	69	12.5	286	2 AE2677	ABC transporter, m
38	69	12.5	286	2 C97459	sulfate ABC transp
39	69	12.5	346	2 C82052	N-acetyl-gamma-glu
40	69	12.5	547	2 H75632	Na(+)-linked D-ala
41	69	12.5	1045	2 D83393	RND multidrug effl
42	68.5	12.4	291	2 B69098	phosphate transport
43	68.5	12.4	311	2 H82541	conserved hypotet
44	68.5	12.4	493	2 A41751	glucose transport
45	68.5	12.4	802	2 C86303	probable vacuolar-

ALIGNMENTS

RESULT 1

D65048
hypothetical protein (emrR 5' region) - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 09-Jul-2004
C:Accession: D65048
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: D65048
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-111 <BLAT>
A:Cross-references: UNIPROT:P43667; UNIPARC:UPI000004F5DE; GB:AE000353; GB:U000096; NID
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: ysaH
C:Superfamily: Escherichia coli hypothetical protein (emrR 5' region)

Query Match 100.0%; Score 552; DB 2; Length 111;
Best Local Similarity 100.0%; Pred. No. 1.5e-49;
Matches 111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSYEVLGLLVGVNYCFRYPRLRVGNARPTKRGAVGILLDTIGIASICALLVVSTA 60
|||
Db 1 MSYEVLGLLVGVNYCFRYPRLRVGNARPTKRGAVGILLDTIGIASICALLVVSTA 60

Qy 61 PEVMDTRRPVTLVGFVAVLGASFYKTRSIITPLLSALAYGLAWKVMII 111
|||
Db 61 PEVMDTRRPVTLVGFVAVLGASFYKTRSIITPLLSALAYGLAWKVMII 111

RESULT 2

A91072
hypothetical protein ECs3545 [imported] - Escherichia coli (strain O157:H7, substrain R
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C:Accession: A91072
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gaawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: A91072
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-111 <HAY>
A:Cross-references: UNIPROT:Q8X907; UNIPARC:UPI0000000E4F0; GB:BA0000007;
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:

A:Gene: ECa3545
C:Superfamily: Escherichia coli hypothetical protein (emrR 5' region)

Query Match 99.3%; Score 548; DB 2; Length 111;
Best Local Similarity 99.1%; Pred. No. 3.9e-49;
Matches 110; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MSYEVLLGLLVGVNYCFRYLPLRLRVGNARPTKRGAVGILLDTIGTIGASTICALLVSTA 60
Db 1 MSYEVLLGLLVGAANYCFRYLPLRLRVGNARPTKRGAVGILLDTIGTIGASTICALLVSTA 60

Qy 61 PEVMDHTRFRFVPTLVGFVAVLGASFYKTRSIITPTLLSALAYGLAWKWMII 111
Db 61 PEVMDHTRFRFVPTLVGFVAVLGASFYKTRSIITPTLLSALAYGLAWKWMII 111

RESULT 3
D85916
hypothetical protein ygaH [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C:Accession: D85916
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouasis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: D85916
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-111 <STO>
A:Cross-references: UNIPROT:Q8X907; UNIPARC:UPI0000165BDF; GB:AE005174; NID:gl2517124; F
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: ygaH
C:Superfamily: Escherichia coli hypothetical protein (emrR 5' region)

Query Match 98.6%; Score 544; DB 2; Length 111;
Best Local Similarity 98.2%; Pred. No. 1e-48;
Matches 109; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MSYEVLLGLLVGVNYCFRYLPLRLRVGNARPTKRGAVGILLDTIGTIGASTICALLVSTA 60
Db 1 MSYEVLLGLLVGAANYCFRYLPLRLRVGNARPTKRGAVGILLDTIGTIGASTICALLVSTA 60

Qy 61 PEVMDHTRFRFVPTLVGFVAVLGASFYKTRSIITPTLLSALAYGLAWKWMII 111
Db 61 PEVMDHTRFRFVPTLVGFVAVLGASFYKTRSIITPTLLSALAYGLAWKWMII 111

RESULT 4
AG0396
probable membrane protein YPO3265 [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004
C:Accession: AG0396
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarragher, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrrell,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AG0396
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-113 <KUR>
A:Cross-references: UNIPROT:Q8ZBW9; UNIPARC:UPI00000DC73C; GB:AL590842; PIDN:CAC92499.1;
C:Genetics:
A:Gene: YPO3265
C:Superfamily: Escherichia coli hypothetical protein (emrR 5' region)

Query Match 55.9%; Score 308.5; DB 2; Length 113;
Best Local Similarity 54.4%; Pred. No. 1.4e-24;

Matches 62; Conservative 21; Mismatches 26; Indels 5; Gaps 2;

Qy 1 MSYEVLLGLLVGVNYCFRYLPLRLRVGNARPTK---RGAVGILLDTIGTIGASTICALLV 57
Db 1 MNTDILMIGLVGVNYCFRYLPLRLRVGNARPTK---GPAKQTGLLRGKMSLLDSIGIASICALLV 58

Qy 58 STAPEVMDHTRFRFVPTLVGFVAVLGASFYKTRSIITPTLLSALAYGLAWKWMII 111
Db 59 SSTPEIHHFQKLLPTLIGLVICGCFYKTRSIITATLLGALCYGLTFKLLTIL 112

RESULT 5
A95272
hypothetical protein SMA0151 [imported] - Sinorhizobium meliloti (strain 1021) magaplasma
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 09-Jul-2004
C:Accession: A95272
R:Barnett, M.J.; Fisher, R.P.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bows
.: Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
A:Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium melilot
A:Reference number: A95262; MUID:21396509; PMID:11481432
A:Accession: A95272
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-866 <KUR>
A:Cross-references: UNIPROT:Q930W3; UNIPARC:UPI000000CAF89; GB:AE006469; PIDN:AAK64739.1;
A:Experimental source: strain 1021, megaplasmid pSymA
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.P.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
heBault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.;
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: SMA0151
A:Genome: plasmid

Query Match 14.2%; Score 78.5; DB 2; Length 866;
Best Local Similarity 25.4%; Pred. No. 3.7;
Matches 30; Conservative 17; Mismatches 40; Indels 31; Gaps 4;

Qy 13 GVANYCFRYLPLRLRVGNA-----RPTKRGAVGILLDTIGTIGASI----- 51
Db 4 GIANLAYVSIPIPMFVLMGAAYVASSPAGSDLYTSLDRWLNRIFGGLILSNIGACAIIFSGMTG 63

Qy 52 -----CALLVVSTAPEVMDHTRFRFVPTLVGFVAVLGASFYKTRSIITPTLLSALAYGLA 104
Db 64 SSPATCAIGKMGIPENM---RRGYPASVASGSIAG--GTLGILIPPSVTLIVYGIA 116

RESULT 6
G87022
hypothetical protein murE [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 05-Oct-2004
C:Accession: G87022
R:Cole, S.T.; Biglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho;
cam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq
A:Title: Massive gene decay in the leprosy bacillus.
A:Reference number: A86909; MUID:21128732; PMID:11234002
A:Accession: G87022
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-530 <STO>
A:Cross-references: UNIPROT:O69557; UNIPARC:UPI000012F9A0; GB:AL450390; NID:gl3092979; P
C:Genetics:

A:Molecule type: DNA
A:Residues: 1-486 <STO>
A:Cross-references: UNIPROT:Q9ZVHS; UNIPARC:UPI00000A210E; GB:AE002093; NID:G3786021; PT:
C:Genetics:
A:Gene: At2g38510
A:Map position: 2

Query Match 13.5% Score 74.5; DB 2; Length 486;
Best Local Similarity 26.9%; Pred.No 5.4;
Matches 28; Conservative 15; Mismatches 30; Indels 31; Gaps 4;

Qy 3 YEVLL--GLLVGVNYCFRYLPLRLVGNARPTKRGAVGILLDTIGIASICALLVVSTA 60
 |||::||| :|::|||:|::|||:|::|||:|::|||:|::|||:|
Db 250 YEIMFLFCGLL-----GNPKASV-AAMGLIQTTLGILYVFPFAISSA- 290

Qy 61 PEVMHDTRRFVPTLVGFANVLGSFYKTRSIITLLSALAYGLA 104
 :|||::|||:|::|||:|::|||:|::|||:|::|||:|::|||:|::|||:|::|||:|
Db 291 -----IATRVGHALGGQPTRAQTTVIGLILAVAYGLA 324

RESULT 11
H70836
hypothetical protein Rv0290 - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004
C:Accession: H70836
R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, K.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: H70836
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-472 <COL>
A:Cross-references: UNIPROT:O8362; UNIPARC:UPI00000D0F99; GB:AL021930; GB:AL123456; NID:
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: Rv0290

Query Match	13.4%	Score 74;	DB 2;	Length 472;
Best Local Similarity	29.5%;	Pred. No. 6;		
Matches	33;	Conservative 13;	Mismatches 36;	Indels 30; Gaps 5;
Qy	7	LLGLLV-----GVANYCFRYLPLRLVGNVARTKGVAILDITGIA--SICALLVV	57	
Db	170	LAGLLITRSPRGTSIALAVPIGAILALAVPGKGPQVLGAAGVAWSLIALLMP	229	
Qy	58	STAPEVMDHTRFVPTLVGFVAVLGASPYKTRSIITPLLSALAYG--LANKV	107	
Db	230	SAERE-----RVVAFFTAAGVAVGASV-----ALAGAQLLQOL	262	

RESULT 12

A49806
prv43 protein - suid herpesvirus 1
C:Species: suid herpesvirus 1
C:Date: 01-Dec-1995 #sequence_revision 01-Dec-1995 #text_change 09-Jul-2004
C:Accession: A49806; S27926
R:Powers, L.; Wilkinson, K.S.; Ryan, P.
Virology 199, 81-88, 1994
A:Title: Characterization of the prv43 gene of pseudorabies virus and demonstration that
A:Reference number: A49806; MUID:94160583; PMID:8116257
A:Accession: A49806
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-373 <POW>
A:Cross-references: UNIPROT:Q69342; UNIPARC:UPI000000F8FAF; EMBL:M77761; NID:G334073; PMID:
A:Note: the source is given as pseudorabies virus
C:Genetics:
A:Gene: prv43

C;Superfamily: varicella-zoster virus gene 15 protein
C;Keywords: early protein

Query Match 13.3%; Score 73.5; DB 2; Length 373;
Best Local Similarity 25.9%; Pred. No. 5.3;
Matches 28; Conservative 21; Mismatches 36; Indels 23; Gaps 6;

Qy 5 VLLGLLGVNVCFRYLPURLRVGNARPKRGAVGILLDTGTIGASICALLVVSTAPEVM 64
Db :||:||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
27 ITLMGVLLAARCLMSVP-----PAASTML--LGVASLMAWL---RLFMPL 68

Qy 65 HDTRFRFVPTLVGPAVLGASPYKTRSIIIPLLLSALAVG-LAWKVMAII 111
Db | ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
69 VD--RFAPACMGQLVGAAVFAAGWALASR--DAISAGULLWAVCALI 112

RESULT 13
AGL1126
Conserved membrane protein [imported] - Listeria monocytogenes (strain EGD-e)
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C;Accession: AGL1126
R;Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, H.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Matheis, S.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend, O.; C;Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend, O.;
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; PMID:21537279; PMID:11679669
A;Accession: AGL1126
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-269 <GLA>
A;Cross-references: UNIPROT:Q8Y9V6; UNIPARC:UPI00000CF0C4; GB_NC_003210; PIDN:CAC98493.1
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: lmo0414
C;Superfamily: gufa protein

	Query Match	13.2%	Score 73;	DB 2;	Length 269;
	Best Local Similarity	33.3%;	Pred. No. 4.;	Mismatches 13;	Gaps 6;
	Matches 35; Conservative				
Qy	2 SYEVLLGLLAVG-	-ANCFRYLPFLRLRVGNARPTKRGAVGILLDTGTIASIC	S2		
Dd	7 SLNPFVLLLAGIFTWACTAAGASLVFFFKLNKKKNGNI--MLGFPAAGVML----	AASF	60		
Qy	53 ALLVVSTAPEVMHMDTRR--FVPVLTVGVFAVLGSAFYKRTSRIIIPTL	I95			
Dd	61 SLYI-APATMSNDLGKFSEFPALVGF-LGGIFLRVIDRIIPHL	I102			

RESULT 14
AD1487
conserved membrane protein [imported] - *Listeria innocua* (strain Clip11262)
C:Species: *Listeria innocua*
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C:Accession: AD1487
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
S.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Feigl, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kraft, J.; Kuhn, M.; Kunst, F.; Kurapkut, G.; Madueno, E.; Maitournam, A.; Mo
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland
A:Title: Comparative genomics of *Listeria* species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AD1487
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-269 <GLA>
A:Cross-references: UNIPROT:Q92EM3, UNIPARC:UPI000000CC223; GB:AL592022; PIDN:CAC95668.1
A:Experimental source: strain Clip11262

C;Genetics:
A;Gene: lin0435
C;Superfamily: gufA protein

Query Match 13.2%; Score 73; DB 2; Length 269;
Best Local Similarity 33.7%; Pred. No. 4.3; Mismatches 20; Gaps 6;
Matches 34; Conservative 13; Indels 34; Indels 20; Gaps 6;
Qy 6 LLLGLLVGV-----ANYCFRYLRLRLVGNARPTKRGAVGILLDTIGIASICALLY 56
Db 11 VLLALLAGIPTWACTAAGASIVFFFKLNKKWGNV--MLGFAAGVWL-----AASFWSLL- 63
Qy 57 VSTAPEVMDTR--FVPTLVGPAVLGASFYKTRSIITPL 95
Db 64 -APAIEMSKDLGKFSFVPAWGF-LLGGIFLRVIDRIIPL 102

RESULT 15

S28025
light harvesting complex II assembly factor pucC [imported] - Rhodobacter sphaeroides
C;Species: Rhodobacter sphaeroides
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 09-Jul-2004
C;Accession: S28025; T50701
R;Gibson, L.C.D.; McGlynn, P.; Chaudhri, M.; Hunter, C.N.
Mol. Microbiol. 6, 3171-3186, 1992
A;Title: A putative anaerobic coproporphyrinogen III oxidase in Rhodobacter sphaeroides.
A;Reference number: S28023; MUID:93086425; PMID:1453956
A;Accession: S28025
A;Molecule type: DNA
A;Residues: 1-459 <GIB>
A;Cross-references: UNIPROT:Q02443; UNIPARC:UPI00001329F0; EMBL:X68796; NID:G46438; PIDN
R;Choudhary, M.; Kaplan, S.
Nucleic Acids Res. 28, 862-867, 2000
A;Title: DNA sequence analysis of the photosynthesis region of Rhodobacter sphaeroides 2
A;Reference number: Z25222; MUID:20115911; PMID:10648776
A;Accession: T50701
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-459 <CHO>
A;Cross-references: UNIPARC:UPI00001329F0; EMBL:AF195122; PIDN:AAF24245.1
A;Experimental source: strain 2.4.1
C;Genetics:
A;Gene: pucC

Query Match 13.2%; Score 73; DB 2; Length 459;
Best Local Similarity 29.0%; Pred. No. 7.3;
Matches 31; Conservative 11; Mismatches 51; Indels 14; Gaps 3;
Qy 5 VLLLG-LLVGAVNYCFRYLRLRLVGNARPTKRGAVGILLDTIGIASICALLYVSTAPEV 63
Db 352 VFLLGTLVVGFGGLFSGHTL---TATWRLAPKEQVGLGAWGAVQATAAGVAIAGAV 408
Qy 64 MHDTRFVPTLVG-----PAVLGASFYKTRSIITPLSALA 100
Db 409 LRDIQAMPDLGSGYGPAPYVAVFAEGFLFLTMIVILPLRSALA 455

Search completed: February 15, 2006, 12:24:27
Job time : 16.9663 secs

This Page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 12:16:07 ; Search time 91.9803 Seconds
(without alignments)
851.417 Million cell updates/sec

Title: US-10-073-293A-6
Perfect score: 552
Sequence: 1 MSFEVLLGLLVGVNYCFR.....IPTLLSALAYGLAWKVMAII 111

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot_05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	552	100.0	111	1 YGAH_ECOLI	P43667 escherichia
2	548	99.3	111	2 Q7ABD5_ECO57	Q7abds escherichia
3	548	99.3	111	2 Q8FEQ5_ECOL6	Q8fed5 escherichia
4	548	99.3	111	2 Q83JZ5_SHIFL	Q83jz5 shigella fl
5	544	98.6	111	2 Q8X907_ECO57	Q8x907 escherichia
6	313.5	56.8	113	2 Q93KB3_ERWCH	Q93kb3 erwinia chr
7	309.5	56.1	113	2 Q66E35_YERPS	Q66e35 yerinia ps
8	309.5	56.1	113	2 Q6D1D8_ERWCT	Q6dl88 erwinia car
9	308.5	55.9	113	2 Q8ZBW9_YERPE	Q8zbw9 yerinia pe
10	271.5	49.2	112	2 Q7N780_PHOLL	Q7n780 photorhabdi
11	87	15.8	408	2 Q6MH96_BDEBA	Q6mh96 bdellovibri
12	86.5	15.7	983	2 Q4NUH1_DEHLT	Q4nuh1 anaeromyxob
13	86	15.6	463	2 Q9HIG1_THEAC	Q9hig1 thermoplasm
14	86	15.6	478	2 Q4V7J8_XENLA	Q4v7j8 xenopus lae
15	83.5	15.1	169	2 Q5TAJ8_HUMAN	O5taj8 homo sapien
16	82.5	14.9	167	2 Q9N3B7_CAEBL	Q9n3b7 caenorhabdi
17	82	14.9	575	2 Q98S82_RHILO	Q98s82 rhizobium l
18	82	14.9	587	2 Q7UQM5_RHOBA	Q7uqm5 rhodospirell
19	82	14.9	589	1 FBX24_MOUSE	Q9d417 mus muscullu
20	81.5	14.8	297	2 Q982L1_RHILO	Q982l1 rhizobium l
21	80.5	14.6	400	1 GTR3_RABIT	Q9xsc2 oryctolagus
22	79.5	14.4	355	2 Q8H78_METAC	Q8h78 methanoearc
23	79	14.3	366	2 Q83G83_TROWT	Q83g83 tropheryma
24	79	14.3	366	2 Q83HY7_TROW8	Q83hy7 tropheryma
25	78.5	14.2	238	2 Q61B17_CAEBR	Q61b17 caenorhabdi
26	78.5	14.2	495	1 GTR3_CANFA	P47842 canis famil
27	78.5	14.2	516	2 Q8EH66_SHEON	Q8eh66 shewanella
28	78.5	14.2	611	2 Q4HVS1_GIBZE	Q4hvs1 gibberella
29	78.5	14.2	866	2 Q930W3_RHIME	Q930w3 rhizobium m
30	78	14.1	530	1 MURE_MYCLE	Q69557 mycobacteri
31	78	14.1	538	2 Q5YTQ8_NOCPA	Q5ytq8 nocardia fa

32	77.5	14.0	365	2	Q872P7_NEUCR	Q872p7 neurospora
33	77.5	14.0	437	2	Q4KFY4_PSEF5	Q4kfy4 pseudomonas
34	77.5	14.0	496	1	GTR3_PONPY	Q5r608 pongo pygma
35	77	13.9	138	2	Q72CW5_DESVH	Q72cw5 desulfovibr
36	77	13.9	358	2	Q7SA63_NEUCR	Q7sa63 neurospora
37	77	13.9	535	2	Q8RVG5_DAUCA	Q8rvgs daucus caro
38	76.5	13.9	309	1	YQGH_BACSU	P46339 bacillus su
39	76.5	13.9	433	2	Q5AWM6_EMENI	Q5awm6 aspergillus
40	76.5	13.9	494	1	GTR3_BOVIN	P58352 bos taurus
41	76.5	13.9	494	1	GTR3_SHEEP	P47843 ovis aries
42	76	13.8	475	2	Q8ABF2_BACTN	Q8abf2 bacteroides
43	76	13.8	502	2	Q93Z26_ARATH	Q93z26 arabidopsis
44	76	13.8	528	2	Q4UW01_XANCP	Q4uw01 xanthomonas
45	76	13.8	528	2	Q8P842_XANCP	Q8p842 xanthomonas

ALIGNMENTS

RESULT 1

ID	YGAH_ECOLI	STANDARD;	PRT;	111 AA.
AC	P43667; P77323;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	10-MAY-2005 (Rel. 47, Last annotation update)			
DE	Hypothetical protein YGAH.			
GN	Names: ygaH; Ordered: locusNames-b2683;			
OS	Escherichia coli.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Escherichia.			
OX	NCBI_TaxID=562;			
RN	[1]			
RP	NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].			
RC	STRAIN=K12 / MG1655;			
RC	MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;			
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,			
RA	Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.P.,			
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,			
RA	Mau B., Shao Y.;			
RT	"The complete genome sequence of Escherichia coli K-12.";			
RL	Science 277:1453-1474 (1997).			
RN	[2]			
RP	NUCLEOTIDE SEQUENCE.			
RC	STRAIN=K12;			
RC	MEDLINE=97349980; PubMed=9205837;			
RA	Yamanoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,			
RA	Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,			
RA	Mizobuchi K., Mori H., Nakade S., Nakamura Y., Washimoto H.,			
RA	Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubram S.,			
RA	Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,			
RA	Yamagata S., Horiuchi T.;			
RT	"Construction of a contiguous 874-kb sequence of the Escherichia coli-			
RT	K12 genome corresponding to 50.0-68.8 min on the linkage map and			
RT	analysis of its sequence features.";			
RL	DNA Res. 4:91-113 (1997).			
RN	[3]			
RP	NUCLEOTIDE SEQUENCE OF 39-111.			
RC	STRAIN=K12;			
RC	MEDLINE=95247664; PubMed=7730261;			
RA	Lomovskaya O., Lewis K., Matin A.;			
RT	"EmrR is a negative regulator of the Escherichia coli multidrug			
RT	resistance pump EmrAB.";			
RL	J. Bacteriol. 177:2328-2334 (1995).			
RN	[4]			
RP	IDENTIFICATION.			
RA	Rudd K.E.;			
CC	Unpublished observations (JUL-1995).			

This Swiss-Prot entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use as long as its content is in no way modified and this statement is not

```
CC removed.
CC -----
DR EMBL; U00096; AAC75730.1; -; Genomic DNA.
DR EMBL; D90891; BAA1545.1; -; Genomic DNA.
DR EMBL; U19993; -; NOT_ANNOTATED_CDS; Genomic DNA.
DR PIR; D65048; D65048.1; -; Genomic DNA.
DR ECHOBASE; EB2775; -; Genomic DNA.
DR EcoGene; EG12940; ygaH.
KW Complete proteome; ygaH.
FT CONFLICT 48 49 IA -> MP (in Ref. 3).
FT CONFLICT 54 54 L -> Q (in Ref. 3).
FT CONFLICT 89 111 SIIPTLLSALAYGLAWKVMII -> THYHPNTA (in Ref. 3).
SQ SEQUENCE 111 AA; 12024 MW; 2527D05A3E1DA69D CRC64;

Query Match 100.0%; Score 552; DB 1; Length 111;
Best Local Similarity 100.0%; Pred. No. 1.5e-44;
Matches 111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSYEVLLGLLVGVANVCFRYLPLRLVGNARPTKRGAVGILLDTIGIASICALLVVSTA 60
Db 1 MSYEVLLGLLVGVANVCFRYLPLRLVGNARPTKRGAVGILLDTIGIASICALLVVSTA 60

Qy 61 PEVMDTRFRVPTLVGFAVLGASFYKTRSIIPITLLSALAYGLAWKVMII 111
Db 61 PEVMDTRFRVPTLVGFAVLGASFYKTRSIIPITLLSALAYGLAWKVMII 111

RESULT 2
Q7ABD5_ECO57 PRELIMINARY; PRT; 111 AA.
ID Q7ABD5;
AC Q7ABD5;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein ECs3545.
GN OrderedLocusNames=ECs3545;
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=O157:H7 / Sakai / RIMD 0509952 / EHEC;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba H., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
DR EMBL; BA000007; BAB36968.1; -; Genomic DNA.
KW Hypothetical protein.
SQ SEQUENCE 111 AA; 11995 MW; 3E392E54DEF3A863 CRC64;

Query Match 99.3%; Score 548; DB 2; Length 111;
Best Local Similarity 99.1%; Pred. No. 3.7e-44;
Matches 110; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MSYEVLLGLLVGVANVCFRYLPLRLVGNARPTKRGAVGILLDTIGIASICALLVVSTA 60
Db 1 MSYEVLLGLLVGVANVCFRYLPLRLVGNARPTKRGAVGILLDTIGIASICALLVVSTA 60

Qy 61 PEVMDTRFRVPTLVGFAVLGASFYKTRSIIPITLLSALAYGLAWKVMII 111
Db 61 PEVMDTRFRVPTLVGFAVLGASFYKTRSIIPITLLSALAYGLAWKVMII 111

RESULT 3
Q8FEQ5_ECOL6 PRELIMINARY; PRT; 111 AA.
ID Q8FEQ5;
AC Q8FEQ5;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Hypothetical protein ygaH.
GN Name=ygaH; OrderedLocusNames=c3236;
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157; DOI=10.1073/pnas.252529799;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Rosech P.,
RA Raeko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli."
PL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AE016765; AA881688.1; -; Genomic DNA.
DR PIR; A91072; A91072.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 111 AA; 11995 MW; 3E392E54DEF3A863 CRC64;

Query Match 99.3%; Score 548; DB 2; Length 111;
Best Local Similarity 99.1%; Pred. No. 3.7e-44;
Matches 110; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MSYEVLLGLLVGVANVCFRYLPLRLVGNARPTKRGAVGILLDTIGIASICALLVVSTA 60
Db 1 MSYEVLLGLLVGVANVCFRYLPLRLVGNARPTKRGAVGILLDTIGIASICALLVVSTA 60

Qy 61 PEVMDTRFRVPTLVGFAVLGASFYKTRSIIPITLLSALAYGLAWKVMII 111
Db 61 PEVMDTRFRVPTLVGFAVLGASFYKTRSIIPITLLSALAYGLAWKVMII 111

RESULT 4
Q83JZ5_SHIFL PRELIMINARY; PRT; 111 AA.
ID Q83JZ5;
AC Q83JZ5;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Hypothetical protein ygaH.
GN Name=ygaH; OrderedLocusNames=S2897, SF2710;
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=301 / Serotype 2a;
RX MEDLINE=22272406; PubMed=12384590; DOI=10.1093/nar/gkf566;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157."
RL Nucleic Acids Res. 30:4432-4441(2002).
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=22590274; PubMed=12704152;
DOI=10.1128/IAI.71.5.2775-2786.2003;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
```



```

RT flexneri serotype 2a strain 2457T. ";
RL Infect. Immun. 71:2775-2786 (2003).
DR EMBL; AE005674; AAN44203.1; -; Genomic DNA.
DR EMBL; AE016987; AAP18030.1; -; Genomic DNA.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 111 AA; 11995 MW; 38392E54DEF3A863 CRC64;

Query Match 99.3%; Score 548; DB 2; Length 111;
Best Local Similarity 99.1%; Pred. No. 3.7e-44;
Matches 110; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MSYEVLGGLGVGNYCFRYLPRLRVGNARPTKRGAVGILLDTIGIASICALLVYSTA 60
Db 1 MSYEVLGGLGVGNYCFRYLPRLRVGNARPTKRGAVGILLDTIGIASICALLVYSTA 60
Qy 61 PEVMDTRFRFVPTLVGFAVLGASFYKTRSIIPITLLSALAYGLAWKVMII 111
Db 61 PEVMDTRFRFVPTLVGFAVLGASFYKTRSIIPITLLSALAYGLAWKVMII 111

RESULT 5
Q8X907 ECOS7
ID Q8X907 ECO57 PRELIMINARY; PRT; 111 AA.
AC Q8X907;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein ygaH.
GN Name=ygaH; OrderedLocusNames=z3984;
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=O157:H7 / EDL933 / ATCC 700927 / EHEC;
RX MEDLINE=21074935; PubMed=11206551; DOI=10.1038/35054089;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533 (2001).
DR EMBL; AE005174; AAC57792.1; -; Genomic DNA.
DR PIR; A91072; A91072.
DR PIR; D85916; D85916.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 111 AA; 12018 MW; 3E392E54DEF3A184 CRC64;

Query Match 98.6%; Score 544; DB 2; Length 111;
Best Local Similarity 98.2%; Pred. No. 8.8e-44;
Matches 109; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MSYEVLGGLGVGNYCFRYLPRLRVGNARPTKRGAVGILLDTIGIASICALLVYSTA 60
Db 1 MSYEVLGGLGVGNYCFRYLPRLRVGNARPTKRGAVGILLDTIGIASICALLVYSTA 60
Qy 61 PEVMDTRFRFVPTLVGFAVLGASFYKTRSIIPITLLSALAYGLAWKVMII 111
Db 61 PEVMDTRFRFVPTLVGFAVLGASFYKTRSIIPITLLSALAYGLAWKVMII 111

RESULT 6
Q93KB3 ERWCH
ID Q93KB3 ERWCH PRELIMINARY; PRT; 113 AA.
AC Q93KB3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE YgaH protein.
GN Name=ygaH;

OS Erwinia chrysanthemi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pectobacterium.
OX NCBI_TaxID=556;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=3937;
RA Reverchon S.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ410307; CAC43438.1; -; Genomic DNA.
SQ SEQUENCE 113 AA; 12338 MW; CC18A5FF9C9523D CRC64;

Query Match 56.8%; Score 313.5; DB 2; Length 113;
Best Local Similarity 58.9%; Pred. No. 6.2e-22;
Matches 63; Conservative 15; Mismatches 28; Indels 1; Gaps 1;

Qy 1 MSYEVLGGLGVGNYCFRYLPRLRVGNARPTKRGAVGILLDTIGIASICALLVYST 59
Db 1 MNTSVLLGLGVGNYCFRYLPRLRVGNARPTKRGAVGILLDTIGIASICALLVYST 59
Qy 60 APEVMDTRFRFVPTLVGFAVLGASFYKTRSIIPITLLSALAYGLAWK 106
Db 61 LPDQMHTKXMTLAGFALLTLCFYKTRSIIVLSTLLGALCYGIVFK 107

RESULT 7
Q66E35 YERPS
ID Q66E35 YERPS PRELIMINARY; PRT; 113 AA.
AC Q66E35;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Putative LIV-E family branched chain amino acid exporter, small
DE subunit.
GN Name=ygaH; OrderedLocusNames=YPTB0858;
GN Yersinia pseudotuberculosis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=633;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=IP32953 / Serotype I;
RX PubMed=15358858; DOI=10.1073/pnas.0404012101;
RA Chain P.S.G., Carniel E., Larimer F.W., Lamerdin J., Stoutland P.O.,
RA Regala W.M., Georgescu A.M., Vergez L.M., Land M.L., Motin V.L.,
RA Brubaker R.R., Fowler J., Hinnebusch J., Marceau M., Medigue C.,
RA Simonet M., Chenal-Francisque V., Souza B., Dacheux D., Elliott J.M.,
RA Derbise A., Hauser L.J., Garcia E.;
RT "Insights into the evolution of Yersinia pestis through whole-genome
RT comparison with Yersinia pseudotuberculosis.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:13826-13831 (2004).
DR EMBL; BX936398; CAH20098.1; -; Genomic DNA.
KW Complete proteome.
SQ SEQUENCE 113 AA; 12300 MW; F244135984358C8A CRC64;

Query Match 56.1%; Score 309.5; DB 2; Length 113;
Best Local Similarity 55.3%; Pred. No. 1.5e-21;
Matches 63; Conservative 20; Mismatches 26; Indels 5; Gaps 2;

Qy 1 MSYEVLGGLGVGNYCFRYLPRLRVGNARPTK--RGAVGILLDTIGIASICALLV 57
Db 1 MNTDVLMIGLVGVGNYCFRYLPRLRVGNARPTK--GPARKQTGLRGKMSLLDSIGIASICALLV 58
Qy 58 STAPEVMDTRFRFVPTLVGFAVLGASFYKTRSIIPITLLSALAYGLAWKVMII 111
Db 59 SSTPEIIHPQKLLPTLIGLVICGCFYKTNISIFATLIGALCYGTFKLLTIL 112

RESULT 8
Q6D1D8 ERWCT
ID Q6D1D8 ERWCT PRELIMINARY; PRT; 113 AA.
AC Q6D1D8;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)

```

```
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DE 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Putative membrane protein.
GN OrderedLocusNames=ECA3510;
OS Erwinia carotovora (subsp. atroseptica) (Pectobacterium atrosepticum).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pectobacterium.
OX NCBI_TaxID=29471,
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=SCRI 1043 / ATCC BAA-672;
RX PubMed=15263089; DOI=10.1073/pnas.0402424101;
RA Bell K.S., Sebahia M., Pritchard L., Holden M.T.G., Hyman L.J.,
RA Holleva M.C., Thomson N.R., Bentley S.D., Churcher L.J.C., Mungall K.,
RA Atkin R., Bason N., Brooks K., Chillingworth T., Clark K., Doggett J.,
RA Fraser A., Hance Z., Hauser H., Jagels K., Moule S., Norbertczak H.,
RA Ormond D., Price C., Quail M.A., Sanders M., Walker D., Whitehead S.,
RA Salmond G.P.C., Birch P.R.J., Parkhill J., Toth I.K.;
RA "Genome sequence of the enterobacterial phytopathogen Erwinia
RT carotovora subsp. atroseptica and characterization of virulence
RT factors.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:11105-11110(2004).
DR EMBL; BX950851; CAG76408.1; -; Genomic_DNA.
KW Complete proteome.
SQ SEQUENCE 113 AA; 12232 MW; EAA747D0E5BA18E7 CRC64;

Query Match 56.1%; Score 309.5; DB 2; Length 113;
Best Local Similarity 56.0%; Pred. No. 1.5e-21;
Matches 61; Conservative 21; Mismatches 26; Indels 1; Gaps 1;

Qy 1 MSYEVLLGLLVGVNYCFRYLPLRLRVGNARPT--KRGAVGILLDTTIGIASICALLVVST 59
Db 1 MSTEVLLGLLVGVNYCFRYLPLRLRVGNARPT--KRGAVGILLDTTIGIASICALLVVST 60

Qy 60 APEVMHDTRRFVPTLVGFAVLGASFYKTRSIIPITLLSALAYGLAWKVM 108
Db 61 VPDILTTHKLPVLTVGPIITLTACFYKTRSIIVSLTLLGAFYCGIAFKLL 109

RESULT 9
ID Q8ZBW9 YERPE PRELIMINARY; PRT; 113 AA.
AC Q8ZBW9 Q74X00; Q7CK79;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE Putative membrane protein (Hypothetical protein y0924).
GN OrderedLocusNames=YPO667, YPO3265, Y0924;
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360; DOI=10.1038/35097083;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C.M., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.-M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M.A., Rutherford K.M.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=KIM5 / Biovar Mediaevalis;
RX MEDLINE=22137863; PubMed=12142430;
RX DOI=10.1126/JB.184.16.4601-4611.2002;
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
```

```
RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
RA Perry R.D.;
RT "Genome sequence of Yersinia pestis KIM.";
RL J. Bacteriol. 184:4601-4611(2002).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=91001;
RX PubMed=15368893;
RA Song Y., Tong Z., Wang J., Wang L., Guo Z., Han Y., Zhang J., Pei D.,
RA Zhou D., Qin H., Pang X., Han Y., Zhai J., Li M., Cui B., Qi Z.,
RA Jin L., Dai R., Chen F., Li S., Ye C., Du Z., Lin W., Wang J., Yu J.,
RA Yang H., Wang J., Huang P., Yang R.;
RT "Complete genome sequence of Yersinia pestis strain 91001, an isolate
RT avirulent to humans.";
RL DNA Res. 11:179-197(2004);
DR EMBL; AJ414156; CAC92499.1; -; Genomic_DNA.
DR EMBL; AE013695; AAM84506.1; -; Genomic_DNA.
DR EMBL; AE017129; AAS60934.1; -; Genomic_DNA.
DR PIR; AG0396; AG0396.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 113 AA; 12314 MW; 1FF2DF284D0406A2 CRC64;

Query Match 55.9%; Score 308.5; DB 2; Length 113;
Best Local Similarity 54.4%; Pred. No. 1.9e-21;
Matches 62; Conservative 21; Mismatches 26; Indels 5; Gaps 2;

Qy 1 MSYEVLLGLLVGVNYCFRYLPLRLRVGNARPTK---RGAVGILLDTTIGIASICALLVV 57
Db 1 MNTDILMIGLVGVNYCFRYLPLRLRVGNARPTK---RGAVGILLDTTIGIASICALLVV 58

Qy 58 STAPEVMHDTRRFVPTLVGFAVLGASFYKTRSIIPITLLSALAYGLAWKVM 111
Db 59 SSTPEIHHKLPVLTVGPIITLTACFYKTRSIIVSLTLLGAFYCGIAFKLL 112

RESULT 10
ID Q7N780 PHOLL PRELIMINARY; PRT; 112 AA.
AC Q7N780;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Similar to unknown protein YgaH of Escherichia coli.
GN OrderedLocusNames=plu1278;
OS Photorhabdus luminescens (subsp. laumondii).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photorhabdus.
OX NCBI_TaxID=141679;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=TT01;
RX MEDLINE=22957627; PubMed=14528314; DOI=10.1038/nbt886;
RA Duchaud E., Rusniok C., Frangeul L., Buchrieser C., Givaudan A.,
RA Taurit S., Bocs S., Bouraux-Eude C., Chandler M., Charles J.-F.,
RA Dassa E., Derose R., Derzelle S., Freysinet G., Gaudriault S.,
RA Medigue C., Lanois A., Powell K., Signier P., Vincent R., Wingate V.,
RA Zouine M., Glaser P., Boenare N., Danchin A., Kunst F.;
RT "The genome sequence of the entomopathogenic bacterium Photorhabdus
RT luminescens.";
RL Nat. Biotechnol. 21:1307-1313(2003).
RX EMBL; BX571863; CAB13572.1; -; Genomic_DNA.
DR Photolista; plu1278; -.
KW Complete proteome.
SQ SEQUENCE 112 AA; 12245 MW; 570FE0F8E73247F8 CRC64;

Query Match 49.2%; Score 271.5; DB 2; Length 112;
Best Local Similarity 50.5%; Pred. No. 5.9e-18;
Matches 54; Conservative 23; Mismatches 25; Indels 5; Gaps 2;

Qy 4 EVLLGLLVGVNYCFRYLPLRLRVGNARPT---KRGAVGILLDTTIGIASICALLVVSTA 60
Db 5 KILLIGLVGLNFSFRLPLRVLPF--GKAROSAGRKAGKTSIILDSIGIASICSLIVSGV 62
```

QY 61 PEVMDTRFRFVPTLVGFVAVLGASFYKTRSIIPITLLSALAYGLAWKV 107
 Db 63 PDVNRSEQKLPITLIGCLTICLVFYKTKQILATFGALLFGLTFKI 109

RESULT 11
 Q6MH96_BDEBA PRELIMINARY; PRT; 408 AA.
 AC Q6MH96;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE MFS permease.
 GN Name-yder; OrderedLocusNames=Bd3656;
 OS Bdellovibrio bacteriovorus.
 OC Bacteria; Proteobacteria; Deltaproteobacteria; Bdellovibrionales;
 OC Bdellovibrionaceae; Bdellovibrio.
 OX NCBI_TaxID=959;
 RN [1]

NUCLEOTIDE SEQUENCE.
 RP STRAIN=HD100 / DSM 50701 / ATCC 15356 / NCIB 9529;
 RC PubMed=14752164; DOI=10.1126/science.1093027;
 RX Rendulic S., Jactap P., Rosinus A., Eppinger M., Baar C., Lanz C.,
 RA Keller H., Lambert C., Evans K.J., Goessmann A., Meyer F.,
 RA Sockett R.E., Schuster S.C.;
 RT "A predator unmasked: life cycle of Bdellovibrio bacteriovorus from a
 genomic perspective.";
 RL Science 303:689-692(2004).
 DR EMBL; BX842656; CAE81031.1; -; Genomic_DNA.
 DR GO; GO:0019866; C:inner membrane; IEA.
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0005215; F:transporter activity; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR007114; MFS.
 DR InterPro; IPR011701; MFS.1.
 DR InterPro; IPR002029; Pept_88_S53.
 DR Pfam; PF07690; MFS_1; 1.
 DR PROSITE; PS00850; MFS; 1.
 DR PROSITE; PS00136; SUBTILASE_ASP; UNKNOWN_1.
 KW Complete proteome.
 SQ SEQUENCE 408 AA; 43528 MW; A353E3898A344610 CRC64;

Query Match 15.8%; Score 87; DB 2; Length 408;
 Best Local Similarity 28.4%; Pred. No. 5.6;
 Matches 31; Conservative 20; Mismatches 54; Indels 4; Gaps

QY 5 VLLGLLVG--VANYCFRYLPRLRVGNAPTKRGAVGILLDTIGTASICALLVSTAPE 62
 Db 27 VLLMAVGIIVANLYAQ-PITAMISQALGDPSSAAGLVVTLTQIGYGLVLLVPLGD 85

QY 63 VMHTRFRFVPTLVGFVAVLGASFYKTRSIIPITLLSALAYGLAWKV 111
 Db 86 II-ENRRLVLTIGIAGLVGLGAPASQLTPFYFAFAFGIGASTVQIL 133

RESULT 12
 Q4NUH1_9DELTA PRELIMINARY; PRT; 983 AA.
 ID Q4NUH1_9DELTA
 AC Q4NUH1;
 DT 13-SEP-2005 (TrEMBLrel. 31, Created)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
 DE Spermine synthase precursor.
 GN ORFNames=AdehDRAFT_2712;
 OS Anaeromyxobacter dehalogenans 2CP-C.
 OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
 OC Cyctobacterineae; Myxococcaceae; Anaeromyxobacter.
 OX NCBI_TaxID=290397;
 RN [1]

NUCLEOTIDE SEQUENCE.
 RP STRAIN=2CP-C;
 RC US DOE Joint Genome Institute (JGI-PGF);
 RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 12:23:52 ; Search time 20.8904 Seconds
(without alignments)
439.292 Million cell updates/sec

Title: US-10-073-293A-6
Perfect score: 552
Sequence: 1 MSYEVLGLLVGVNYCFR.....IPTLLSALAYGLAWKVAII 111

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572050 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/6_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/H_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/PCRTUS_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/RE_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	403	73.0	123	2	US-09-489-039A-10218
2	182.5	33.1	122	2	US-09-328-352-6864
3	76.5	13.9	494	1	US-09-031-392-5
4	76.5	13.9	494	2	US-09-299-549-5
5	76.5	13.9	494	2	US-09-610-417-5
6	76.5	13.9	494	2	US-09-981-947B-5
7	76	13.8	655	2	US-09-134-001C-5531
8	74	13.4	202	2	US-09-603-208A-228
9	74	13.4	202	2	US-09-603-208A-230
10	73.5	13.3	332	2	US-09-489-039A-9330
11	72.5	13.1	404	2	US-09-543-681A-6702
12	72.5	13.1	496	2	US-10-146-704-3
13	72	13.0	438	2	US-09-543-681A-8247
14	71.5	13.0	479	2	US-09-605-703B-1560
15	70	12.7	294	2	US-09-328-352-7553
16	70	12.7	429	2	US-09-328-352-4392
17	70	12.7	502	2	US-09-248-796A-16824
18	69.5	12.6	371	2	US-09-252-991A-20081
19	69.5	12.6	459	2	US-09-252-991A-22668
20	69	12.5	534	2	US-09-328-352-6713
21	69	12.5	566	2	US-09-248-796A-20411
22	68.5	12.4	251	2	US-09-758-759-87
23	68	12.3	424	2	US-09-107-532A-5579
24	68	12.3	219	2	US-09-252-991A-23245
25	68	12.3	711	2	US-09-711-164-456
26	67.5	12.2	217	2	US-09-902-540-15768
27	67.5	12.2	355	2	US-09-489-039A-12579

28	67.5	12.2	467	2	US-09-540-236-2469	Sequence 2469, Ap
29	67.5	12.2	1040	2	US-09-902-540-11105	Sequence 11105, A
30	67	12.1	339	2	US-09-489-039A-7277	Sequence 7277, Ap
31	67	12.1	487	2	US-09-489-039A-9810	Sequence 9810, Ap
32	67	12.1	733	2	US-09-489-039A-11108	Sequence 11108, A
33	66.5	12.0	177	2	US-09-710-279-2532	Sequence 2532, Ap
34	66.5	12.0	292	2	US-09-408-020-72	Sequence 72, Appl
35	66.5	12.0	350	2	US-09-252-991A-32358	Sequence 32358, A
36	66.5	12.0	372	2	US-09-328-352-6169	Sequence 6169, Ap
37	66.5	12.0	428	2	US-08-311-731A-380	Sequence 380, App
38	66.5	12.0	492	2	US-09-328-352-6875	Sequence 6875, Ap
39	66.5	12.0	815	2	US-09-902-540-10502	Sequence 10502, A
40	66.5	12.0	1055	2	US-09-902-540-11135	Sequence 11135, A
41	66	12.0	447	2	US-09-489-039A-20825	Sequence 20825, A
42	66	12.0	1040	2	US-09-489-039A-13619	Sequence 13619, A
43	65.5	11.9	288	2	US-09-252-991A-25627	Sequence 25627, A
44	65.5	11.9	509	1	US-09-031-392-6	Sequence 6, Appli
45	65.5	11.9	509	2	US-09-299-549-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-09-489-039A-10218
; Sequence 10218, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 10218
; LENGTH: 123
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-10218

Query Match 73.0%; Score 403; DB 2; Length 123;
Best Local Similarity 72.0%; Pred. No. 4.3e+40;
Matches 77; Conservative 13; Mismatches 17; Indels 0; Gaps 0;
Qy 1 MSYEVLGLLVGVNYCFRYPRLRVGNVARTKRGAVGILLDTGTGICALLVVSTA 60
Db 13 MNSEVLLGVVGVNYLFRLRLRAGHARPARPGVSVLLDTGTGICALLVVSSV 72
Qy 61 PEVMDTRFRVPTLVGVFAVLGASFYKTRSIITPLLSALAYGLAWKV 107
Db 73 PEILADAHRLAPTLTGLVGAFAFWKTRSLIVPTLLSAFAYLAWKI 119

RESULT 2
US-09-328-352-6864
; Sequence 6864, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6864
; LENGTH: 122
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6864

```

; APPLICANT: Tartaglia, Louis A.
; APPLICANT: Weng, Xun
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: GLUTEX AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/299,549
; FILING DATE: 26-APR-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/031,392
; FILING DATE: 26-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Melklejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 07334/072002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 494 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-299-549-5

Query Match 13.9%; Score 76.5; DB 2; Length 494;
Best Local Similarity 34.1%; Pred.No.1.1;
Matches 29; Conservative 17; Mismatches 30; Indels 9; Gaps 7;

QY 2 SYEVLLIG-LLGVG-ANYCFRYLPLRLRVGNARPTK-RGAVGILLDTIGIASICALLVVS 58
Db 116 SVEMLILGRLLIGLFCGLCTGFPV--MYIGEISPTALRGAFG-TLNQLGI--VIGILVAQ 170
QY 59 T-APENVHDTRRFVPTLVGFAVLGA 82
Db 171 IFGLKVLGTEDLWPLLLGFTILPA 195

RESULT 5
US-09-610-417-5
; Sequence 5, Application US/09610417
; Patent No. 6346374
; GENERAL INFORMATION:
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; GLUTEX AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95

```

```

; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/610.417
; FILING DATE: 05-Jul-2000
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/299,549
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Meikiejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 07334/072002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 494 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-610-417-5

```

```

Query Match      13.9%; Score 76.5; DB 2; Length 494;
Best Local Similarity 34.1%; Pred. No. 1.1;
Matches 29; Conservative 17; Mismatches 30; Indels 9; Gaps 7;

QY  2 SYEVLLG-LLGV-ANYCFRYLRLRVGNARPTK-RGAVGILLDTGTIGIASICALVWS 58
Db  116 SVEMLILGRLLIGLFCGLCTGFVP--MYIGISPTALRGAGF-TLNQLGI--VIGILVAQ 170
QY  59 T-APEVMHDTRRFVPTLVGFAVLGA 82
Db  171 IFGLKVILGTEDLWPLLGLFTILPA 195

```

```

RESULT 6
US-09-981-947B-5
; Sequence 5, Application US/09981947B
; Patent No. 6933115
; GENERAL INFORMATION:
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding GLUTX
; TITLE OF INVENTION: and Uses Thereof
; FILE REFERENCE: MPI1998-021DV3
; CURRENT APPLICATION NUMBER: US/09/981,947B
; CURRENT FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 09/610,417
; PRIOR FILING DATE: 2000-07-05
; PRIOR APPLICATION NUMBER: 09/299,549
; PRIOR FILING DATE: 1999-04-26
; PRIOR APPLICATION NUMBER: 09/031,392
; PRIOR FILING DATE: 1998-02-26
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 494
; TYPE: PRT
; ORGANISM: human
US-09-981-947B-5

```

```

Query Match      13.9%; Score 76.5; DB 2; Length 494;
Best Local Similarity 34.1%; Pred. No. 1.1;
Matches 29; Conservative 17; Mismatches 30; Indels 9; Gaps 7;

QY  2 SYEVLLG-LLGV-ANYCFRYLRLRVGNARPTK-RGAVGILLDTGTIGIASICALVWS 58
Db  116 SVEMLILGRLLIGLFCGLCTGFVP--MYIGISPTALRGAGF-TLNQLGI--VIGILVAQ 170
QY  59 T-APEVMHDTRRFVPTLVGFAVLGA 82

```

```

Db  171 IFGLKVILGTEDLWPLLGLFTILPA 195

RESULT 7
US-09-134-001C-5531
; Sequence 5531, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5531
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5531

```

```

Query Match      13.8%; Score 76; DB 2; Length 655;
Best Local Similarity 21.0%; Pred. No. 1.8;
Matches 34; Conservative 21; Mismatches 39; Indels 68; Gaps 7;

QY  5 VLLGLLVGV-----ANYCFRYLRLRVGNARPTKRGAVGILLDTGTIGIASIC 52
Db  479 IMILGLVIGANMAIDMGPFNKAAVVF--ATAALTEGNAAPITAAMIGMIPPLAIAT-- 534
QY  53 ALLAVV-----STAPEVMHDTRRFVPT-LVG----- 76
Db  535 AMLIFRRKTYKRGSGIVPNVYMGSLSPITEGALPFFAAADPLRVIPSMVGVGATALG 594
QY  77 -----PAVLGASF-----YKTRSIITPLLSALAYG 102
Db  595 LGSSIKAPHGGIFVIIGTDFNHLIQTLLALVGLTVLSALIYG 636

```

```

RESULT 8
US-09-603-208A-228
; Sequence 228, Application US/09603208A
; Patent No. 6822084
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberhauer, Gregor
; APPLICANT: Lee, Heung-Shick
; APPLICANT: Kim, Hyung-Joon
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING STRESS,
; FILE REFERENCE: BGI-124CP
; CURRENT APPLICATION NUMBER: US/09/603,208A
; CURRENT FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: 60/141031
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: 60/142692
; PRIOR FILING DATE: 1999-07-01
; PRIOR APPLICATION NUMBER: 60/151214
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: DE 19930429.7
; PRIOR FILING DATE: 1999-07-01
; PRIOR APPLICATION NUMBER: DE 19931413.6
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931457.8
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931541.8
; PRIOR FILING DATE: 1999-07-08

```


Matches 32; Conservative 19; Mismatches 35; Indels 23; Gaps 7;
Qy 6 LLLGLLVG--VANTCFRYPLRLRVG---NARPT--KRGAVG-----ILLDTIGIASIC 52
Db 169 LVGLGIGPVARILIRNIPFP-GAGSDHEMPTAFKPGTGRMITSWVLETTAMISIC 227
Qy 53 ALLVVSTAPEVMDTRFVPTLVGAVLGASFYKTRSIITPLLSALAY 101
Db 228 -LWAGTSLQLLEGTAFSLPTFVCLVFIG-----VILNSLSMLGF 267
RESULT 12
US-10-146-704-3
; Sequence 3, Application US/10146704
; Patent No. 6828421
; GENERAL INFORMATION:
; APPLICANT: Myriad Genetics, Incorporated
; APPLICANT: Sugiyama, Janice
; TITLE OF INVENTION: VAMP-ASSOCIATED PROTEIN A-INTERACTING PROTEINS AND USE THEREOF
; FILE REFERENCE: 1669.01
; CURRENT APPLICATION NUMBER: US/10/146,704
; CURRENT FILING DATE: 2004-05-14
; PRIOR APPLICATION NUMBER: US 60/291,730
; PRIOR FILING DATE: 2001-05-17
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 496
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-146-704-3
Query Match 13.1%; Score 72.5; DB 2; Length 496;
Best Local Similarity 32.9%; Pred. No. 3.4;
Matches 28; Conservative 17; Mismatches 31; Indels 9; Gaps 7;
Qy 2 SYEVLLLG-LLVG-ANYCFRYPLRLRVGNARPTK-RGAVGILLDTIGIASICALVVS 58
Db 116 SVEMILIGRLVIGLFCGLCTGFVP--MYIGISPTALRGAFG-TLNQLGI--VVGILVAQ 170
Qy 59 T-APVMDTRFVPTLVGFAVLGA 82
Db 171 IFGLEFILGSEELWPLLGLFTILPA 195

RESULT 13
US-09-543-681A-8247
; Sequence 8247, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 8247
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-8247
Query Match 13.0%; Score 72; DB 2; Length 438;
Best Local Similarity 25.2%; Pred. No. 3.3;
Matches 31; Conservative 24; Mismatches 40; Indels 28; Gaps 6;
Qy 1 MSBEVLLGLLVGVANTCFRYPLRLRVGNARPTKRG-----AVGILLDTIGI--- 48
Db 90 LGVELALGGFIVCGVLFC-----LVAGIVKTAGRGWVNPFPFPAAMGAIVAVIGLELA 142

Qy 49 ---ASICALLVSTAPEVMDTRFVPTLV--GFAVLGASFYKTRSIITPLLSALA-YG 102
Db 143 GTAANMAGLLPSADNPV---DSQTLIISMVTLGVTILGSMFRGLAIIPILIGLAGYA 199
Qy 103 LAW 105
Db 200 LSF 202
RESULT 14
US-09-605-703B-1560
; Sequence 1560, Application US/09605703B
; Patent No. 6962989
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberer, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING NOVEL
; FILE REFERENCE: BGI-129CP
; CURRENT APPLICATION NUMBER: US/09/605,703B
; CURRENT FILING DATE: 2000-06-27
; PRIOR APPLICATION NUMBER: 60/142,764
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: 60/152,318
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 2934
; SEQ ID NO 1560
; LENGTH: 479
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-605-703B-1560
Query Match 13.0%; Score 71.5; DB 2; Length 479;
Best Local Similarity 27.3%; Pred. No. 4.2;
Matches 39; Conservative 16; Mismatches 45; Indels 43; Gaps 7;
Qy 4 EVLLGLLVGV-----ANY-----CFRYPLRLRVGNARPTKRG----- 37
Db 16 ELFVKGLTGEGTWTLTNIGANYIGFPPLLTVPILLAVGVAE--RSGMLAALIRKLF 73
Qy 38 -----AVGILLDTIGIASICALLVSTAPEVMDHT--RRFVPTLVG-FAVLGASFY 85
Db 74 SAKKIVLPYAVGVIGVYGTASIMADAFFVVPPLAAMVFAAGRHPVAGLLGSFAAGVAGY- 132
Qy 86 KTRSIITPLLSALAYGLAWKVM 108
Db 133 --STAIVPTSLDALFAGITNAV 153

RESULT 15
US-09-328-352-7553
; Sequence 7553, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7553
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7553
Query Match 12.7%; Score 70; DB 2; Length 294;
Best Local Similarity 25.8%; Pred. No. 3.5;
Matches 33; Conservative 19; Mismatches 48; Indels 28; Gaps 5;

Qy	1	MSYEVLLGLLVGVANYCFRY-----LPLRLRVGNARPTRKGVAGILLDDTTGISICAL	54
Db	26	MLYELFLGLLSGVTTWLFGGGGFVAVPLTYIIQKWSNBSSVGIHAMQIAVATSAFV	85
Qy	55	LWVSTAPVMDTR-----RPVPTLVGFVAVLGASFYKTRSIIPTLLSALAYG--L	103
Db	86	MLCSASFVAFRHYRSGHIDWQKIRE---LWGGIALGG-----IVGAVMASLFGNGWL	134
Qy	104	AWKVMAIL 111	
Db	135	RWIFMGYV 142	

Search completed: February 15, 2006, 12:25:39
Job time : 21.8904 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.
OM protein - protein search, using sw model
Run on: February 15, 2006, 12:24:42 ; Search time 71.4017 Seconds
(without alignments)
649.551 Million cell updates/sec
Title: US-10-073-293A-6
Perfect score: 552
Sequence: 1 MSYEVLLGLLVGVNYCFR.....IPTLLSALAYGLAWKVMII 111
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 1867569 seqs, 417829326 residues
Total number of hits satisfying chosen parameters: 1867569
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : Published Applications AA Main:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	552	100.0	111	5	US-10-073-293A-6
2	84	15.2	223	4	US-10-424-599-223037
3	80.5	14.6	400	5	US-10-732-923-23642
4	79.5	14.4	1238	4	US-10-424-599-161652
5	78.5	14.2	495	5	US-10-732-923-23556
6	78	14.1	530	4	US-10-282-122A-63783
7	77.5	14.0	442	4	US-10-282-122A-62944
8	77	13.9	308	4	US-10-369-493-3529
9	76.5	13.9	161	5	US-10-732-923-23831
10	76.5	13.9	494	3	US-09-981-947A-5
11	76.5	13.9	494	5	US-10-732-923-23560
12	76.5	13.9	494	5	US-10-732-923-23639
13	76.5	13.9	500	4	US-10-369-493-12445
14	76	13.8	509	4	US-10-092-900A-314
15	76	13.8	538	4	US-10-437-963-107875
16	76	13.8	650	4	US-10-282-122A-70668
17	76	13.8	655	4	US-10-724-972A-7255
18	75.5	13.7	201	5	US-10-501-282-5130
19	75.5	13.7	225	5	US-10-501-282-5132
20	75.5	13.7	253	5	US-10-501-282-5134
21	75.5	13.7	254	5	US-10-501-282-5136
22	75.5	13.7	283	5	US-10-501-282-5138
23	75.5	13.7	326	5	US-10-501-282-5140
24	75.5	13.7	327	5	US-10-501-282-5142
25	75.5	13.7	362	4	US-10-437-963-115286
26	75.5	13.7	497	5	US-10-732-923-23714
27	75.5	13.7	520	5	US-10-732-923-23713

28	74.5	13.5	551	5	US-10-968-848-84	Sequence 84, Appl
29	74	13.4	250	4	US-10-425-115-215825	Sequence 215825,
30	74	13.4	472	4	US-10-080-170-380	Sequence 380, App
31	74	13.4	472	4	US-10-080-170-380	Sequence 380, App
32	74	13.4	472	4	US-10-458-356-380	Sequence 380, App
33	74	13.4	791	3	US-09-738-626-4530	Sequence 4530, Ap
34	73.5	13.3	156	5	US-10-852-335A-111	Sequence 111, App
35	72.5	13.1	496	4	US-10-170-385-247	Sequence 247, App
36	72.5	13.1	496	4	US-10-341-434-89	Sequence 89, Appl
37	72.5	13.1	496	4	US-10-755-889-666	Sequence 666, App
38	72.5	13.1	496	5	US-10-893-315-85	Sequence 85, Appl
39	72.5	13.1	499	5	US-10-893-315-92	Sequence 92, Appl
40	72	13.0	185	4	US-10-425-115-342079	Sequence 342079,
41	72	13.0	481	4	US-10-437-963-199115	Sequence 199115,
42	71.5	13.0	225	4	US-10-282-122A-53684	Sequence 53684, A
43	71.5	13.0	488	5	US-10-732-923-11192	Sequence 11192, A
44	71.5	13.0	489	5	US-10-732-923-23545	Sequence 23545, A
45	71.5	13.0	502	4	US-10-156-761-12709	Sequence 12709, A

ALIGNMENTS

RESULT 1

US-10-073-293A-6
; Sequence 6, Application US/10073293A
; Publication No. US20050239175A1
; GENERAL INFORMATION:
; APPLICANT: TABOLINA, EKATERINA
; APPLICANT: RYBAK, KONSTANTIN
; APPLICANT: KHOURGES, EVGENI
; APPLICANT: VOROSHILOVA, ELVIRA
; APPLICANT: GUSYATINER, MIKHAIL
; TITLE OF INVENTION: METHOD FOR PRODUCING L-AMINO ACID USING BACTERIA BELONGING TO TI
; TITLE OF INVENTION: ESCHERICHIA
; FILE REFERENCE: 219594USO
; CURRENT APPLICATION NUMBER: US/10/073,293A
; CURRENT FILING DATE: 2002-02-13
; PRIOR APPLICATION NUMBER: RU 2001103865
; PRIOR FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: RU 2001104998
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: RU 2001104999
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: RU 2001117632
; PRIOR FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: RU 2001117633
; PRIOR FILING DATE: 2001-06-28
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 111
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-073-293A-6

Query Match 100.0%; Score 552; DB 5; Length 111;
Best Local Similarity 100.0%; Pred. No. 3.1e-59; Indels 0; Gaps 0;
Matches 111; Conservative 0; Mismatches 0;
QY 1 MSYEVLLGLLVGVNYCFRYPRLRVGNARPTKRGAVGILLDTIGIASICALLVVSTA 60
Db 1 MSYEVLLGLLVGVNYCFRYPRLRVGNARPTKRGAVGILLDTIGIASICALLVVSTA 60
QY 61 PEVMDTRFPVTLGVFAVLGASFYKTRSIIPITLLSALAYGLAWKVMII 111
Db 61 PEVMDTRFPVTLGVFAVLGASFYKTRSIIPITLLSALAYGLAWKVMII 111

RESULT 2

US-10-424-599-223037
; Sequence 223037, Application US/10424599
; Publication No. US20040031072A1

```
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 223037
; LENGTH: 223
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(400)
; OTHER INFORMATION: Clone ID: PAT_MRT3847_43431C.1.pep
US-10-424-599-223037

Query Match 15.2%; Score 84; DB 4; Length 223;
Best Local Similarity 29.8%; Pred. No. 0.11;
Matches 25; Conservative 16; Mismatches 21; Indels 22; Gaps 4;

Qy 39 VGILLDTIGIASICALLVSTAPEVMDTRRF-----VPTLVGFAVLGA--SFYKTRSI 92
Db 132 VWITDKI-----LMLFSQDPEISHAREYCYILIPALFGHVLQALTRYFQTQSWIP 184

Qy 93 PTLISA-----LAYGLAWK 107
Db 185 PMVFSITALCLHPICWGLVFKL 208

RESULT 3
US-10-732-923-23642
; Sequence 23642, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 23642
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(400)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-732-923-23642

Query Match 14.6%; Score 80.5; DB 5; Length 400;
Best Local Similarity 31.1%; Pred. No. 0.62;
Matches 32; Conservative 18; Mismatches 32; Indels 21; Gaps 8;

Qy 2 SYEVLLLG-LLVGV-ANYCFRPLRLRVGNARPTK-RGAVGILLDTIGIASICALLVVS 58
Db 19 SVMELILGRLLIGFCGLCTGFVP--MVIGESPTALRGAGF-TLNQLGI--VIGILVAQ 73

Qy 59 T-APEVMDTRRFVPTLVGFAVLGASFYKTRSIITPTLSALA 100
Db 74 IFGLEIILGSVLPWPLVLLGFT-----IIPALQSAA 104

RESULT 4
US-10-424-599-161652
; Sequence 161652, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
```

```
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 161652
; LENGTH: 1238
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1238)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_116990C.1.pep
US-10-424-599-161652

Query Match 14.4%; Score 79.5; DB 4; Length 1238;
Best Local Similarity 31.8%; Pred. No. 3.4;
Matches 21; Conservative 13; Mismatches 19; Indels 13; Gaps 3;

Qy 41 ILDDTIGIASICALLVSTAPEVMDTRRF-----VPTLVGFAVLGAS--FYKTRSIIPT 94
Db 96 IFVDKI-----LLLFQDQPEISHAHEYCIYIPALYGLQCFVAVLQCFYFQTQSMIFPM 148

Qy 95 LLSALA 100
Db 149 VFSSIA 154

RESULT 5
US-10-732-923-23556
; Sequence 23556, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 23556
; LENGTH: 495
; TYPE: PRT
; ORGANISM: Canis familiaris
; OTHER INFORMATION: unsure at all Xaa locations
US-10-732-923-23556

Query Match 14.2%; Score 78.5; DB 5; Length 495;
Best Local Similarity 31.1%; Pred. No. 1.4;
Matches 32; Conservative 18; Mismatches 32; Indels 21; Gaps 8;

Qy 2 SYEVLLLG-LLVGV-ANYCFRPLRLRVGNARPTK-RGAVGILLDTIGIASICALLVVS 58
Db 116 SVMELILGRLLIGFCGLCTGFVP--MVIGESPTALRGAGF-TLNQLGI--VIGILVAQ 170

Qy 59 T-APEVMDTRRFVPTLVGFAVLGASFYKTRSIITPTLSALA 100
Db 171 IFGLKVMINGTEELWPLLLGFT-----IIPAVLOSAA 201

RESULT 6
US-10-282-122A-63783
; Sequence 63783, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
```

```

; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 62944
; LENGTH: 442
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-10-282-122A-62944

Query Match 14.0%; Score 77.5; DB 4; Length 442;
Best Local Similarity 24.8%; Pred.No. 1.6;
Matches 32; Conservative 20; Mismatches 36; Indels 41; Gaps 6;

QY 9 GLLGVGVNYCPRY-----LPLRLRVGNARPTKRGAVGILLDTGIASICALL 55
DB 255 GALIGLISACTAFTFANIDLILDADLPMQLLVANVHPW-----LGTILMSFIFA----M 305
QY 56 VVSTAPEVMH-----DTRRFVPTLVGFAVLG-----ASFYKTRSIITPT-----LL 96
DB 306 IFNTAISLYSLAKRFSGDDNQRPXKWLIGLVGVGILSFAGFKLVSIMFPILGVIGWL 365
QY 97 SALAYGLAW 105
DB 366 LLVLLLLAW 374

RESULT 8
US-10-369-493-3529
; Sequence 3529, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 3529
; LENGTH: 308
; TYPE: PRT
; ORGANISM: Neurospora crassa

```

FILED DATE: 18-OCC-2001
PRIOR APPLICATION DATA:

FILE REFERENCE: 38-15(52796)C
CURRENT ADDICTION NUMBER: HS/10/722 922

FILE REFERENCE: 38-15(52796)C

```

; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 23639
; LENGTH: 494
; TYPE: PRT
; ORGANISM: Ovis aries
US-10-732-923-23639

Query Match
Best Local Similarity 34.1%; Score 76.5; DB 5; Length 494;
Matches 29; Conservative 17; Mismatches 30; Indels 9; Gaps 7;

QY 2 SYEVLLG-LLGV-ANYCFRYLRLRVGNARPTK-RGAVGILLDTIGIASICALVWS 58
Db 116 SVEMLILGRILGLFCGLCTGFVP--MYIGISPTALRGAFG-TLNQLGI--VIGILVAQ 170

QY 59 T-APEVMHDTRRFVPTLVGFVILGA 82
Db 171 IFGLKVLGTEDLWPLLLGLFTILPA 195

RESULT 13
US-10-369-493-12445
; Sequence 12445, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 12445
; LENGTH: 500
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)...(500)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-12445

Query Match
Best Local Similarity 25.2%; Score 76.5; DB 4; Length 500;
Matches 33; Conservative 20; Mismatches 45; Indels 33; Gaps 5;

QY 9 GLLVGVANYCFRYLRLRVGNARPTKRGAVGILLDTIGIASICALVWSTAPEVMHDT 68
Db 20 GLFVGIGAYLRDAGPLSLGLGY----LIWGIAFILPILNSVGMCAVLPIRGS--IFELAA 74

QY 69 RFVPTLVGFVILGASFYK---TRS-----IIPTLSALAYGL----- 103
Db 75 RYVDPAPGAFGAMXGPSAPRVQGRSDNIRLGDWVVFYAGLMVLCVCTEYSAVAFIMDYQIDV 134

QY 104 ---AWKYMALII 111
Db 135 NPAAWVAMW 145

RESULT 14
US-10-092-900A-314
; Sequence 314, Application US/10092900A
; Publication No. US20040043382A1
; GENERAL INFORMATION:
```

```

; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Taupier Jr., Raymond J.
; APPLICANT: Pena, Carol E.A.
; APPLICANT: Li, Li
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Ji, Weizhen
; APPLICANT: Gorman, Linda
; APPLICANT: Miller, Charles E.
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Patturajan, Meera
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Guo, Xiaojia Sasha
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Fernandes, Elma R.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Liu, Yi
; APPLICANT: Anderson, David W.
; APPLICANT: Spaderna, Steven K.
; APPLICANT: Catterton, Elina
; APPLICANT: Leite, Mario W.
; APPLICANT: Zhong, Haihong
; APPLICANT: Alsobrook, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: No. US20040043382A1el Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-290C
; CURRENT APPLICATION NUMBER: US/10/092,900A
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: USSN 60/274,322
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: USSN 60/283,675
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: USSN 60/338,092
; PRIOR FILING DATE: 2001-12-03
; PRIOR APPLICATION NUMBER: USSN 60/274,281
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: USSN 60/274,191
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: USSN 60/325,681
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: USSN 60/304,354
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: USSN 60/279,995
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: USSN 60/294,899
; PRIOR FILING DATE: 2001-05-31
; PRIOR APPLICATION NUMBER: USSN 60/287,424
; PRIOR FILING DATE: 2001-04-30
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 768
; SEQ ID NO 314
; LENGTH: 509
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-092-900A-314

Query Match
Best Local Similarity 32.1%; Score 76; DB 4; Length 509;
Matches 27; Conservative 17; Mismatches 34; Indels 6; Gaps 5;

QY 2 SYEVLLG-LLGV-ANYCFRYLRLRVGNARPTK-RGAVGILLDTIGIASICALVWS 58
Db 119 SVEMLILGRILGLFCGLCTGFVP--MYIGISPTALRGAFG-TLNQLGIVIGILVAQVI 175

QY 59 TAPEVMHDTRRFVPTLVGFVILGA 82
```

Db 176 FGLEILGSEELWPVLGFTILPA 199

RESULT 15

US-10-437-963-107875
 ; Sequence 107875, Application US/10437963
 ; Publication No. US20040123343A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa, Thomas J.
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Wu, Wei
 ; APPLICANT: Boukharov, Andrey A.
 ; APPLICANT: Barbazuk, Brad
 ; APPLICANT: Li, Ping
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53221)B
 ; CURRENT APPLICATION NUMBER: US/10/437,963
 ; CURRENT FILING DATE: 2003-05-14
 ; NUMBER OF SEQ ID NOS: 204966
 ; SEQ ID NO 107875
 ; LENGTH: 538
 ; TYPE: PRT
 ; ORGANISM: Oryza sativa
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT4530_12184C.1.pap
 US-10-437-963-107875

Query Match 13.8%; Score 76; DB 4; Length 538;
 Best Local Similarity 26.8%; Pred No. 3.2;
 Matches 30; Conservative 15; Mismatches 27; Indels 40; Gaps 6;
 QY 6 LLLGLLVGVANYCFYLPRLRLRVGNARPTKRGAVGILLD-----TIGIASICAL- 54
 Db 84 LMVGLVGVPSY---YL-----AGSLVDLGMSALQGVATVAFANLVLV 124
 QY 55 -LVVSTAPEVMHDTRRFVPTLVGFVILGASFYKTRSIITLLSLMAYGLAW 105
 Db 125 SLVLTAPAVTHG-----LPFVLARAAFGVGRGAHLPAVIRALV-GCGW 167

Search completed: February 15, 2006, 12:29:34
 Job time : 72.4017 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2006, 12:25:52 ; Search time 6.54775 Seconds
(without alignments)
240.922 Million cell updates/sec

Title: US-10-073-293A-6

Perfect score: 552

Sequence: 1 MSTEVLLGLLVGVNYCFR.....IPTLLSALAYGLAWKNVAIL 111

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 107799 seqs, 14211699 residues

Total number of hits satisfying chosen parameters: 107799

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA New*

- 1: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	74	13.4	202	7	US-11-082-389-364
2	72.5	13.1	499	6	US-10-821-234-1106
3	67	12.1	1034	6	US-10-392-234A-30
4	66.5	12.0	177	6	US-10-793-626-2532
5	66.5	12.0	203	6	US-10-453-372-890
6	64.5	11.7	185	6	US-10-453-372-896
7	64.5	11.7	194	6	US-10-453-372-894
8	64.5	11.7	218	6	US-10-453-372-892
9	64	11.6	272	7	US-11-098-686-10661
10	64	11.6	332	6	US-10-454-437-28
11	64	11.6	1037	6	US-10-392-234A-36
12	63.5	11.5	332	6	US-10-511-538-79
13	63.5	11.5	551	6	US-10-055-877-276
14	63.5	11.5	551	6	US-10-055-877-277
15	63	11.4	1048	6	US-10-392-234A-14
16	63	11.4	1048	6	US-10-392-234A-20
17	62.5	11.3	485	6	US-10-467-657-326
18	62	11.2	314	6	US-10-511-538-70
19	62	11.2	1049	6	US-10-392-234A-12
20	61.5	11.1	434	6	US-10-821-234-1680
21	61.5	11.1	520	7	US-11-098-686-11110
22	61	11.1	325	6	US-10-510-101-156
23	61	11.1	983	6	US-10-511-538-91
24	61	11.1	1048	6	US-10-392-234A-18
25	60.5	11.0	294	6	US-10-793-626-1688

26	60.5	11.0	1046	6	US-10-392-234A-16
27	59.5	10.8	228	7	US-11-210-316-16
28	59.5	10.8	295	6	US-10-858-730-112
29	59.5	10.8	480	7	US-11-080-991-76
30	59.5	10.8	518	6	US-10-821-234-1165
31	59	10.7	244	6	US-10-793-626-134
32	59	10.7	352	6	US-10-793-626-944
33	59	10.7	374	7	US-11-228-364-4
34	59	10.7	654	7	US-11-120-308-52
35	59	10.7	895	7	US-11-098-686-10311
36	58.5	10.6	233	6	US-10-858-730-234
37	58.5	10.6	233	7	US-11-055-822-54
38	58.5	10.6	454	6	US-10-055-877-204
39	58.5	10.6	747	7	US-11-210-316-2
40	58.5	10.6	1061	7	US-11-059-814-18
41	58	10.5	311	6	US-10-980-388-111
42	58	10.5	311	6	US-10-980-388-113
43	58	10.5	530	6	US-10-980-388-62
44	57.5	10.4	303	7	US-11-052-554A-321
45	57.5	10.4	488	6	US-10-055-877-231

ALIGNMENTS

RESULT 1

US-11-082-389-364
; Sequence 364, Application US/11082389
; Publication No. US20050244935A1
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberkauer, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS INVOLVED IN MEMBRANE SYNTHESIS AND MEMBRANE TRANSPORT
; TITLE OF INVENTION: TRANSPORT
; FILE REFERENCE: BGI-131CPCN
; CURRENT APPLICATION NUMBER: US/11/082,389
; CURRENT FILING DATE: 2005-03-16
; PRIOR APPLICATION NUMBER: US 09/603024
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/141031
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 60/143262
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: US 60/151281
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: DE 19930487.4
; PRIOR FILING DATE: 1999-07-01
; PRIOR APPLICATION NUMBER: DE 19930489.0
; PRIOR FILING DATE: 1999-07-01
; PRIOR APPLICATION NUMBER: DE 19931549.3
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931550.7
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19932134.5
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19941379.7
; PRIOR FILING DATE: 1999-08-31
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 446
; SEQ ID NO 364
; LENGTH: 202
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-11-082-389-364

Query Match 13.4%; Score 74; DB 7; Length 202;

Best Local Similarity 27.8%; Pred. NO. 0.43; Mismatches 18; Indels 22; Gaps 5;
Matches 30; Conservative 18; Mismatches 38; Indels 22; Gaps 5;

Db 143 LVTQEFNPTSPVNARYEFGPALFVGWASAGLAVLGGSF 181

RESULT 7.

US-10-453-372-894

Sequence 894, Application US/10453372

Publication No. US20060003323A1

GENERAL INFORMATION:

APPLICANT: Alsobrook, et al.

TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD

FILE REFERENCE: 21402-589 A

CURRENT APPLICATION NUMBER: US/10/453,372

CURRENT FILING DATE: 2003-06-03

PRIOR APPLICATION NUMBER: 09/789390

PRIOR FILING DATE: 2001-02-23

PRIOR APPLICATION NUMBER: 60/185967

PRIOR FILING DATE: 2000-03-01

PRIOR APPLICATION NUMBER: 09/823187

PRIOR FILING DATE: 2001-03-29

PRIOR APPLICATION NUMBER: 60/195792

PRIOR FILING DATE: 2000-03-10

PRIOR APPLICATION NUMBER: 09/839446

PRIOR FILING DATE: 2001-03-19

PRIOR APPLICATION NUMBER: 60/199476

PRIOR FILING DATE: 2000-03-25

PRIOR APPLICATION NUMBER: 09/863776

PRIOR FILING DATE: 2001-05-23

PRIOR APPLICATION NUMBER: 60/208263

PRIOR FILING DATE: 2000-05-31

PRIOR APPLICATION NUMBER: 09/939398

PRIOR FILING DATE: 2001-08-24

PRIOR APPLICATION NUMBER: 60/227800

PRIOR FILING DATE: 2000-08-25

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 1609

SOFTWARE: Curasequelist version 0.1

SEQ ID NO 894

LENGTH: 194

TYPE: PRT

ORGANISM: Homo sapiens

US-10-453-372-894

Query Match 11.7%; Score 64.5; DB 6; Length 194;

Best Local Similarity 27.3%; Pred. No. 4.2;

Matches 27; Conservative 15; Mismatches 38; Indels 19; Gaps 5;

QY 5 VLLGLGLLVGVNVCFRYLPLRLRVGNARPTKRGAVGILLDTIGI-ASICALLVVS---- 58

Db 63 LMTVAVLGLFVAVMLSVWGKMKTRVGDSPNPIAKGRVAIAGGALFILAGLCTLTAVSWYAT 122

QY 59 -----TAEVMDHTR-RFVPTL-----VGFAVLGASF 84

Db 123 LVTQEFNPTSPVNARYEFGPALFVGWASAGLAVLGGSF 161

RESULT 8

US-10-453-372-892

Sequence 892, Application US/10453372

Publication No. US20060003323A1

GENERAL INFORMATION:

APPLICANT: Alsobrook, et al.

TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD

FILE REFERENCE: 21402-589 A

CURRENT APPLICATION NUMBER: US/10/453,372

CURRENT FILING DATE: 2003-06-03

PRIOR APPLICATION NUMBER: 09/789390

PRIOR FILING DATE: 2001-02-23

PRIOR APPLICATION NUMBER: 60/185967

PRIOR FILING DATE: 2000-03-01

PRIOR APPLICATION NUMBER: 09/823187

PRIOR FILING DATE: 2001-03-29

PRIOR APPLICATION NUMBER: 60/195792

PRIOR FILING DATE: 2000-03-10

```

; PRIOR APPLICATION NUMBER: 09/839446
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/199476
; PRIOR FILING DATE: 2000-03-25
; PRIOR APPLICATION NUMBER: 09/863776
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 60/208263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/939398
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/227800
; PRIOR FILING DATE: 2000-08-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1609
; SOFTWARE: Curasequest version 0.1
; SEQ ID NO 892
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-453-372-892

Query Match      11.7%; Score 64.5; DB 6; Length 218;
Best Local Similarity 27.3%; Pred. No. 4.8;
Matches 27; Conservative 15; Mismatches 38; Indels 19; Gaps 5;

Qy  5 VLLGLLVGVNVCYFVPLR-LVGNARPTKRGAVGILLDTIGI-ASICALLVVS----- 58
Db  87 LMMVAVLLGFMVNLVSVVGMKTRVGSNPKRGRVAIAGALFILAGLCTLTAVSWYAT 146
Qy  59 -----TAPEVMHDTIR-FVPTL-----VGFAVLGASF 84
Db  147 LVTQEFNPTPVNARVEFGFALFVGWASAGLAVLGGSF 185

RESULT 9
US-11-098-686-10661
; Sequence 10661, Application US/11098686
; Publication No. US20060024696A1
; GENERAL INFORMATION:
; APPLICANT: Kapur, Vivek and Gebhart, Connie J.
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES
; FILE REFERENCE: 09531-128001
; CURRENT APPLICATION NUMBER: US/11/098,686
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: PCT/US03/31318
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: US 60/416,395
; PRIOR FILING DATE: 2002-10-04
; NUMBER OF SEQ ID NOS: 11433
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10661
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Lawsonia intracellularis
US-11-098-686-10661

Query Match      11.6%; Score 64; DB 7; Length 272;
Best Local Similarity 24.3%; Pred. No. 7;
Matches 28; Conservative 23; Mismatches 48; Indels 16; Gaps 6;

Qy  1 MSYEVLLGLLVGVNVCYFVPLR-LVGNARPTKRGAVGILLDTIGI-ASICALLVV 57
Db  1 MSIELLYTGVLTQGLTATIAFGIMIPFRF-LNFPDLTAEGAY-----PLG-GAVCASLMV 53
Qy  58 STAPEVMHDTIRFVPTLVGFAVLGAS-----FVKTRSIITPTLLGALAYGLAWKVM 108
Db  54 ANMFQILAWMAGIMAG--GLLAVGTQSVALRLKVNLLAGILLSTWYSINLRIM 106

RESULT 10
US-10-454-437-28
; Sequence 28, Application US/10454437

```

```

; Publication No. US20050277115A1
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberhauer, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS
; TITLE OF INVENTION: INVOLVED IN HOMEOSTASIS AND ADAPTATION
; FILE REFERENCE: BGI-128CPCN
; CURRENT APPLICATION NUMBER: US/10/454,437
; CURRENT FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: US 60/141031
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: DE 19931636.8
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19932125.6
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932126.4
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932127.2
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932128.0
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932129.9
; PRIOR FILING DATE: 1999-07-19
; PRIOR APPLICATION NUMBER: DE 19932226.0
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932920.6
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: DE 19932922.2
; PRIOR FILING DATE: 1999-07-14
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 442
; SEQ ID NO 28
; LENGTH: 332
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-10-454-437-28

Query Match      11.6%; Score 64; DB 6; Length 332;
Best Local Similarity 26.1%; Pred. No. 8.7;
Matches 30; Conservative 13; Mismatches 38; Indels 34; Gaps 7;

Qy  31 APPTKRGAVGI-----LLD-TIGIASIC-----ALLVVSTAPEV---MHDTRRRFVPTLV- 75
Db  58 AAPLAAGADAIFNORTIIDVAIAFVFCGASAIYLVNDARDVDREHPTRFRPRTAAG 117
Qy  76 -----GPAV-----LGASFYKTRSIITPTLLSA-----LAYGLAWKVM 111
Db  118 VLPVGMAYGMAVALIALSIGLSFLATDGVALACVIGVYIALQLGCFGKHPVI 172

RESULT 11
US-10-392-234A-36
; Sequence 36, Application US/10392234A
; Publication No. US2005025538A1
; GENERAL INFORMATION:
; APPLICANT: Pharmacia and Upjohn Corporation
; APPLICANT: Buxser, Steven
; APPLICANT: Poole, Keith
; APPLICANT: Decker, Douglas
; APPLICANT: Xiaozhi Li
; TITLE OF INVENTION: Method for Screening for acraB Transporter Family Inhibitors
; FILE REFERENCE: 6206
; CURRENT APPLICATION NUMBER: US/10/392,234A
; CURRENT FILING DATE: 2003-03-17
; PRIOR APPLICATION NUMBER: US 60/364,935
; PRIOR FILING DATE: 2002-03-15
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 36
; LENGTH: 1037

```

```

; APPLICANT: Ballinger, Robert
; APPLICANT: Gerlach, Valerie
; APPLICANT: Spytek, Kimberly
; APPLICANT: Ratelli, Luca
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Guo, Xiaojia
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Andrew, David
; APPLICANT: Mezes, Peter
; APPLICANT: Patturajan, Meera
; APPLICANT: Burgess, Cahterine
; APPLICANT: Eisen, Andrew
; APPLICANT: Wolenc, Adam
; APPLICANT: Baumgartner, Jason
; APPLICANT: Shinkets, Richard
; APPLICANT: Gusev, Vladimir
; APPLICANT: Vernet, Corine
; APPLICANT: Taupier Jr., Raymond
; APPLICANT: Pena, Carol
; APPLICANT: Shenoy, Suresh
; APPLICANT: Li, Li
; APPLICANT: Casman, Stacie
; APPLICANT: Boldog, Ference
; TITLE OF INVENTION: Novel Polypeptides and Nucleic Acids Encoded Thereby
; FILE REFERENCE: 21402-251
; CURRENT APPLICATION NUMBER: US/10/055,877
; CURRENT FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: 60/262,892
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/263,598
; PRIOR FILING DATE: 2001-01-23
; PRIOR APPLICATION NUMBER: 60/263,799
; PRIOR FILING DATE: 2001-01-24
; PRIOR APPLICATION NUMBER: 60/264,117
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 60/264,139
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 60/264,478
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/263,351
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/272,870
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: 60/275,990
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/275,927
; PRIOR FILING DATE: 2001-03-14
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 512
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 276
; LENGTH: 551
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-055-877-276

Query Match 11.5%; Score 63.5; DB 6; Length 551;
Best Local Similarity 25.6%; Pred. No. 17;
Matches 30; Conservative 15; Mismatches 37; Indels 35; Gaps

Qy 4 EYLLAGL-LVGVANYCFRFLP-----LRLRVGNAPRTKRGAVGILLDTIGTASTCALL
Db 163 KVLILNYLQTAVSGTCAAYPNYTVYCVFLLSG-----MSLASIAINCMTL;

Qy 56 VVSTAPEVMHTRFVPTLVGFAVLGASFYKTRSIIPTLLSALAYGLA-KVWMAI 111
Db 210 NVEMWP--IH-TRAYVGTLLIGVYSLGQF-----LLAGIAYAVPHWRHLQIV 253

RESULT 14
US-10-055-877-277
; Sequence 277, Application US/10055877
; Publication No. US20050288241A1

```

Db 210 NVEWMP--IH-TRAYVGTLLGYVYSLQCF-----LLAGIAYAVPHRHQLQV 253

RESULT 15

US-10-392-234A-14

; Sequence 14, Application US/10392234A

; Publication No. US2005025538A1

; GENERAL INFORMATION:

; APPLICANT: Pharmacia and Upjohn Corporation

; APPLICANT: Buxser, Steven

; APPLICANT: Poole, Keith

; APPLICANT: Decker, Douglas

; APPLICANT: Xiazunhi Li

; TITLE OF INVENTION: Method for Screening for acrAB Transporter Family Inhibitors

; FILE REFERENCE: 6206

; CURRENT APPLICATION NUMBER: US/10/392,234A

; CURRENT FILING DATE: 2003-03-17

; PRIOR APPLICATION NUMBER: US 60/364,935

; PRIOR FILING DATE: 2002-03-15

; NUMBER OF SEQ ID NOS: 67

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 14

; LENGTH: 1048

; TYPE: PRT

; ORGANISM: Klebsiella pneumoniae

US-10-392-234A-14

Query Match 11.4%; Score 63; DB 6; Length 1048;

Best Local Similarity 28.6%; Pred. No. 40;

Matches 20; Conservative 17; Mismatches 23; Indels 10; Gaps 2;

QY 42 LLDITIGIASICALLVWSTAPEVMDTRRRFVPTLVGFVGLGASFYKTRSIIPITLLSALAY 101

Db 930 LLTTIGLSAKNAIIVFAXDLMEKE-----GKGLIEATLEAVRMRLRPILMTSLAF 981

QY 102 GLAWKVMAII 111

Db 982 ILG--VMPLV 989

Search completed: February 15, 2006, 12:30:01

Job time : 7.54775 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 16, 2006, 17:05:59 ; Search time 2814.6 Seconds
(without alignments)
2241.749 Million cell updates/sec

Title: US-10-073-293a-6
Perfect score: 552
Sequence: 1 MSYEVLLGLLVGVNYCFR.....IPTLLSALAYGLAWKVMAII 111

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 5883141 seqs, 28421725653 residues
Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/absa/ABSSWEB spool/US10073293/runat_15022006_120734_963/app query fasta.1
-DB=GenEmbl -QPM=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abs04
-USER=US10073293 @CGN 1 1 5142 @runat 15022006_120734_963 -NCPU=6 -ICPU=3
-NO MMAP -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_ov:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_hgt:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	552	100.0	336	6	BD177949 Process f
2	552	100.0	336	6	CS078099 Sequence
3	552	100.0	336	6	CS078115 Sequence

4	552	100.0	336	6	AX534670	AX534670 Sequence
5	552	100.0	19150	1	D90891	D90891 E.coli geno
6	552	100.0	110000	1	U00096	Continuation (28 o
7	552	100.0	110000	1	U00096	Continuation (29 o
8	548	99.3	110000	1	AE005674	Continuation (28 o
9	548	99.3	110000	1	BA000007	Continuation (36 o
10	548	99.3	290380	1	AE016987	AE016987 Shigella
11	548	99.3	305325	1	AE016765	AE016765 Escherich
12	544	98.6	110000	1	AE005174	Continuation (36 o
13	544	98.6	110000	1	AE005174	Continuation (37 o
14	403	73.0	372	6	AR386318	AR386318 Sequence
15	313.5	56.8	3619	1	ECH410307	Au410307 Erwinia c
16	309.5	56.1	110000	1	EX936399	Continuation (11 o
17	309.5	56.1	110000	1	EX950851	Continuation (40 o
18	308.5	55.9	10733	1	AE013695	AE013695 Yersinia
19	308.5	55.9	220050	1	AJ414156	AJ414156 Yersinia
20	308.5	55.9	290002	1	AE017129	AE017129 Yersinia
21	295	53.4	328	1	ECU19993	U19993 Escherichia
22	271.5	49.2	243072	6	AX770904	AX770904 Sequence
23	271.5	49.2	349318	1	EX571863	EX571863 Phototab
24	182.5	33.1	369	6	AR320188	AR320188 Sequence
25	164	29.7	100	6	CQ000312	CQ000312 Sequence
26	160	29.0	100	6	CQ000311	CQ000311 Sequence
27	155	28.1	100	6	CQ000313	CQ000313 Sequence
28	147	26.6	114080	14	AP007521	AP007521 Lotus cor
29	87	15.8	307337	1	EX842656	EX842656 Bdellovib
30	86.5	15.7	90018	14	AC017853	AC017853 Drosophil
31	86.5	15.7	167026	2	AC010011	AC010011 Drosophil
32	86.5	15.7	195757	2	AC010009	AC010009 Drosophil
33	86.5	15.7	249021	2	AE003596	AE003596 Drosophil
34	86	15.6	1851	5	BC097869	BC097869 Xenopus l
35	86	15.6	268356	1	TACID5	AL445067 Thermopla
36	85	15.4	233445	14	AC129682	AC129682 Rattus no
37	84.5	15.3	921	15	AY857266	AY857266 Hypocrea
38	83	15.0	632	15	AF348090	AF348090 Trichoder
39	82.5	14.9	19557	13	AY029400	AY029400 Campolet
40	82	14.9	951	4	AY355763	AY355763 Canis fam
41	82	14.9	110000	1	BA000012	Continuation (63 o
42	82	14.9	110000	14	AP006495	AP006495 Cyanidios
43	82	14.9	303250	1	BX294143	BX294143 Pirellula
44	81.5	14.8	483	4	OAUS9030	U99030 Ovis aries
45	81.5	14.8	1793	4	AY033938	AY033938 Bos tauri

ALIGNMENTS

RESULT 1	BD177949	Process for producing L-amino acid using escherichia.	336 bp	DNA	linear	PAT 16-APR-2003
LOCUS	BD177949					
DEFINITION	BD177949					
ACCESSION	BD177949					
VERSION	BD177949.1	GI:30015212				
KEYWORDS	JP 2002300874-A/4.					
SOURCE	Escherichia coli					
ORGANISM	Escherichia coli					
REFERENCE	1	(bases 1 to 336)				
AUTHORS	Tabolina,E.A., Rybak,K.V., Khourges,E.M., Voroshilova,E.B. and Gulyatiner,M.M.					
TITLE	Process for producing L-amino acid using escherichia					
JOURNAL	Patent: JP 2002300874-A 4 15-OCT-2002;					
COMMENT	OS Escherichia coli					
	PN JP 2002300874-A/4					
	PD 15-OCT-2002					
	PR 13-FEB-2001 RU 2001103865,26-FEB-2001 RU 2001104998 PR					
	26-FEB-2001 RU 2001104999,28-JUN-2001 RU 2001117632 PR					
	28-JUN-2001 RU 2001117633					
	PI EKATERINA ALEKSANDROVNA TABOLINA, KONSTANTIN VYACHESLAVOVICH					
	PI RYBAK,					
	PI EVGENI MOISEVICH KHOURGES, ELVIRA BORISOVNA VOROSHILOVA, PI					

MIKHAIL MAROVICH GUSYATINER
 PC C12N1/21, C12P13/04, C12P13/06, C12P13/08, C12P13/12, PC
 C12P13/24//
 PC (C12N1/21, C12R1:19), (C12P13/04, C12R1:19), (C12P13/06, C12R1:19),
 PC (C12P13/08, C12R1:19), (C12P13/08, C12R1:19), (C12P13/12, C12R1:19), PC
 (C12P13/24, C12R1:19)
 CC Process for producing L-amino acid using *Escherichia FH* Key
 Location/Qualifiers

FT CDS Location/Qualifiers
 (1)..(333).

FEATURES

source
 1..336
 /organism="Escherichia coli"
 /mol_type="genomic DNA"
 /db_xref="taxon:562"

ORIGIN

Alignment Scores:
 Pred. No.: 1.16e-50 Length: 336
 Score: 552.00 Matches: 111
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 100.0% Indels: 0
 DB: 6 Gaps: 0

US-10-073-293A-6 (1-111) x BD177949 (1-336)

Qy 1 MetSerTyrGluValLeuLeuGlyLeuValGlyValAlaAseThrCysPheArg 20
 Db 1 ATGAGCTATGAGGTTCTGCTGGGTACTAGTTGGCGTGGCGAATTATTGCTTCGCG 60
 Qy 21 TyrLeuProLeuArgLeuArgValGlyAsnAlaArgProThrLysArgGlyAlaValGly 40
 Db 61 TATTGGCGCTGGCGCTGGGTGTAATGCCCGCCCAACCAACGTCGCGGTAGGT 120
 Qy 41 IleLeuLeuAspThrIleGlyIleAlaSerIleCysAlaLeuLeuValValSerThrAla 60
 Db 121 ATTTTGTCCACACCATTTGGCATGCCCTCGATATGGCTCTGCTGGTGTCTCTACCGCA 180
 Qy 61 ProGluValMetHisAspThrArgPheValProThrLeuValGlyPheAlaValLeu 80
 Db 181 CCAGAGTGATGACGATACACGCGGTTTCGTGGCCACGCTGCGGTCTGCGGTACTG 240
 Qy 81 GlyAlaSerPheTyrLysThrArgSerIleIleIleProThrLeuLeuSerAlaLeuAla 100
 Db 241 GTGCGCAGTTCTTATAAACACGACGATTCATCCCAACACTGCTTAGTGGCGTGGCC 300

RESULT 2
 CS078099
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Escherichia coli
 Escherichia coli
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Escherichia.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 FEATURES
 source

Tabolina,E.A., Rybak,K.V., Kfourges,E.M., Voroshilova,E.B. and
 Gussyatiner,M.M.
 Method for producing L-amino acid using bacteria belonging to the
 genus *Escherichia*
 Patent: EP 1526179-A 5 27-APR-2005;
 Ajinomoto Co., Inc. (JP)
 Location/Qualifiers
 1..336
 /organism="Escherichia coli"
 /mol_type="unassigned DNA"

/db_xref="taxon:562"
 1..336
 /note="unnamed protein product"

/codon_start=1
 /transl_table=11
 /protein_id="CAI94224.1"
 /db_xref="GI:63093068"
 /translation="MSYEVLILGLLVGVANVCFRYLPLRLRVGNARPTKRGAVGILID
 TIGIASICALLVVSTAPEVMHDTFRFPTLVGFVAVLGASFYKTRSIITPTLLSALAY
 LAWKVMAII"

ORIGIN

Alignment Scores:
 Pred. No.: 1.16e-50 Length: 336
 Score: 552.00 Matches: 111
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 100.0% Indels: 0
 DB: 6 Gaps: 0

US-10-073-293A-6 (1-111) x CS078099 (1-336)

Qy 1 MetSerTyrGluValLeuLeuGlyLeuValGlyValAlaAseThrCysPheArg 20
 Db 1 ATGAGCTATGAGGTTCTGCTGGGTACTAGTTGGCGTGGCGAATTATTGCTTCGCG 60
 Qy 21 TyrLeuProLeuArgLeuArgValGlyAsnAlaArgProThrLysArgGlyAlaValGly 40
 Db 61 TATTGGCGCTGGCGCTGGGTGTAATGCCCGCCCAACCAACGTCGCGGTAGGT 120
 Qy 41 IleLeuLeuAspThrIleGlyIleAlaSerIleCysAlaLeuLeuValValSerThrAla 60
 Db 121 ATTTTGTCCACACCATTTGGCATGCCCTCGATATGGCTCTGCTGGTGTCTCTACCGCA 180
 Qy 61 ProGluValMetHisAspThrArgPheValProThrLeuValGlyPheAlaValLeu 80
 Db 181 CCAGAGTGATGACGATACACGCGGTTTCGTGGCCACGCTGCGGTCTGCGGTACTG 240
 Qy 81 GlyAlaSerPheTyrLysThrArgSerIleIleIleProThrLeuLeuSerAlaLeuAla 100
 Db 241 GTGCGCAGTTCTTATAAACACGACGATTCATCCCAACACTGCTTAGTGGCGTGGCC 300
 Qy 101 TyrGlyLeuAlaTrpLysValMetAlaIleIle 111
 Db 301 TATGGGCTCGCTGGAAAGTGATGGCGATTATA 333

RESULT 3

CS078115
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Escherichia coli
 Escherichia coli
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Escherichia.

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

Location/Qualifiers

1..336

/organism="Escherichia coli"

/mol_type="unassigned DNA"

/db_xref="taxon:562"

1..336

/note="unnamed protein product"

/codon_start=1

/transl_table=11

/protein_id="CAI94228.1"
/db_xref="GI:63093084"
/translation="MSYEVLGLLGVGVANYCFRPLRLRVGNARPTKRGAVGILLD
TIGIASICALLVSTAPEVMDTRRFVTLVGFALVGAFKFKTRSIITLLSALAYG
LAWKVMII"

ORIGIN

Alignment Scores:

Pred. No.: 1.16e-50 Length: 336
Score: 552.00 Matches: 111
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 6 Gaps: 0

US-10-073-293A-6 (1-111) x CS078115 (1-336)

Qy 1 MetSerTyrGluValLeuLeuGlyLeuValGlyValAlaAseTyrCysPheArg 20
Db 1 ATGAGCTATGAGGTTCTGCTGCTGGGTTACTAGTTGGCGTGGCGAATTATTGCTCCGC 60
Qy 21 TyrLeuProLeuArgLeuArgValGlyAsnAlaArgProThrLysArgGlyAlaValGly 40
Db 61 TATTGCGCTGCGCTGCGTGGTAAATGCCCGCCCAACCAACGTCGGCGGTAGGT 120
Qy 41 IleLeuLeuAspThrIleGlyIleAlaSerIleCysAlaLeuLeuValSerThrAla 60
Db 121 ATTTTGTGCACACCAATTGGCATCGCTCGATATGGCTCTGCTGGTGTCTCTACCGCA 180
Qy 61 ProGluValMetHisAspThrArgPheValProThrLeuValGlyPheAlaValLeu 80
Db 181 CCAGAAAGTGATGCACGATACACGCCGTTTCGTGCCACACGCTGGTTCGGGTACTG 240
Qy 81 GlyAlaSerPheTyrLysThrArgSerIleIleIleProThrLeuLeuSerAlaLeuAla 100
Db 241 GTTGCCAGTTCTATAAACAACGACGATTCATCCCAACATGCTTAGTGCGTGGCC 300
Qy 101 TyrGlyLeuAlaTrpLysValMetAlaIleIle 111
Db 301 TATGGCGCTCGCTGAAAGTGATGGCGATTATA 333

RESULT 4

AX534670
LOCUS
DEFINITION
Sequence 5 from Patent EP1239041.
ACCESSION
AX534670
VERSION
AX534670.1 GI:25261076

SOURCE

Escherichia coli
Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.

REFERENCE

1 Tabolina,E.A., Rybak,K.V., Khourges,E.M., Voroshilova,E.B. and Gussyatiner,M.M.

Method for producing l-amino acid using bacteria belonging to the genus escherichia

Patent: EP 1239041-A 5 11-SEP-2002;

Ajinomoto Co., Inc. (JP)

FEATURES

source

1. .336
Location/Qualifiers
/organism="Escherichia coli"
/mol_type="unassigned DNA"
/db_xref="taxon:562"

CDS

1. .336
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAD57226.1"
/db_xref="GI:25261077"
/translations="MSYEVLGLLGVGVANYCFRPLRLRVGNARPTKRGAVGILLD
TIGIASICALLVSTAPEVMDTRRFVTLVGFALVGAFKFKTRSIITLLSALAYG
LAWKVMII"

ORIGIN

Alignment Scores:

Pred. No.: 1.16e-50 Length: 336
Score: 552.00 Matches: 111
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 6 Gaps: 0

US-10-073-293A-6 (1-111) x AX534670 (1-336)

Qy 1 MetSerTyrGluValLeuLeuGlyLeuValGlyValAlaAseTyrCysPheArg 20
Db 1 ATGAGCTATGAGGTTCTGCTGCTGGGTTACTAGTTGGCGTGGCGAATTATTGCTCCGC 60
Qy 21 TyrLeuProLeuArgLeuArgValGlyAsnAlaArgProThrLysArgGlyAlaValGly 40
Db 61 TATTGCGCTGCGCTGCGTGGTAAATGCCCGCCCAACCAACGTCGGCGGTAGGT 120
Qy 41 IleLeuLeuAspThrIleGlyIleAlaSerIleCysAlaLeuLeuValSerThrAla 60
Db 121 ATTTTGTGCACACCAATTGGCATCGCTCGATATGGCTCTGCTGGTGTCTCTACCGCA 180
Qy 61 ProGluValMetHisAspThrArgPheValProThrLeuValGlyPheAlaValLeu 80
Db 181 CCAGAAAGTGATGCACGATACACGCCGTTTCGTGCCACACGCTGGTTCGGGTACTG 240
Qy 81 GlyAlaSerPheTyrLysThrArgSerIleIleIleProThrLeuLeuSerAlaLeuAla 100
Db 241 GTTGCCAGTTCTATAAACAACGACGATTCATCCCAACATGCTTAGTGCGTGGCC 300
Qy 101 TyrGlyLeuAlaTrpLysValMetAlaIleIle 111
Db 301 TATGGCGCTCGCTGAAAGTGATGGCGATTATA 333

RESULT 5

D90891
LOCUS
DEFINITION
E.coli genomic DNA, Kohara clone #445(2.0-60.6 min.).
ACCESSION
D90891
VERSION
D90891.1 GI:1800054

KEYWORDS

Complete and shotgun sequencing; HUZY; MG230; NRDE; ZK632.10; emrA;

emrB; emrR; gshA; gshI; hnsB; mprA; nrdF; proU; proV; proW; proX;

stdA; ygaC; ygaG; ygaH; yjdJ; yzzM.

Escherichia coli

Escherichia coli

Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

Enterobacteriaceae; Escherichia.

1 (sites)

Yamamoto,Y., Aiba,H., Baba,T., Hayashi,K., Inada,T., Isono,K.,

Itoh,T., Kimura,S., Kitagawa,M., Makino,K., Miki,T., Mitsuhashi,N.,

Mizobuchi,K., Mori,H., Nakade,S., Nakamura,Y., Nashimoto,H.,

Oshima,T., Oyama,S., Saito,N., Sampei,G., Satoh,Y.,

Sivasubramam,S., Tagami,H., Takahashi,H., Takeda,J., Takemoto,K.,

Uehara,K., Wada,C., Yamagata,S. and Horiuchi,T.

Construction of a contiguous 874-kb sequence of the Escherichia

coli -K12 genome corresponding to 50.0-68.8 min on the linkage map

and analysis of its sequence features

DNA Res. 4 (2), 91-113 (1997)

2 (sites)

Aiba,H., Baba,T., Fujita,K., Hayaehi,K., Honjo,A., Horiuchi,T.,

Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K.,

Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M.,

Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H.,

Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T.,

Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,

Yamamoto,Y. and Yano,M.

The systematic sequencing of the Escherichia coli genome in Japan

Unpublished

REFERENCE
3 (bases 1 to 19150)

Mori,H.

TITLE JOURNAL	Direct Submission	
	Submitted (23-JAN-1997) Hirotada Mori, NARA Institute of Science and Technology, Res. & Edu. Center for Genetic Info.; 8916-5 Takayama, Ikoma, Nara 630-01, Japan (E-mail: hmori@gtc.aist-nara.ac.jp, Tel:81-7437-2-5660, Fax:81-7437-2-5669) Collaboration Information: Project: The Japan E.coli genome DNA sequencing project	
COMMENT	Group: The Japan E.coli genome DNA sequencing group Members: (1995.4 - 1996.3) Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Itohchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S., Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nishimoto,H., Nishio,Y., Oshima,T., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C., Yamamoto,Y. and Yano,M. Headed by: Name: Takashi Horiuchi Address: National Institute of Basic Biology, Okazaki, 444, Japan E-mail: kishori@nibb.ac.jp Information operator: Name: Hirotada Mori Address: NARA Institute of Science and Technology, Ikoma, 630-01, Japan E-mail: hmori@gtc.aist-nara.ac.jp URL: The Japan E. coli genome database http:bsw3.aist-nara.ac.jp. Location/Qualifiers 1. .19150 /organism="Escherichia coli" /mol_type="genomic DNA" /strain="K12" /map="60.2 min" /db_xref="taxon:562" /clone="Kohara clone #445" /clone_lib="Kohara lambda miniset library" /note="Nucleotide position 2791307-2810456 from the initiation site of ThrA (0 min). This clone is from Kohara lambda miniset library" complement(1. .253) /gene="yzzM" complement(<1. .>253) /gene="yzzM" /note="similar to [SwissProt Accession Number P39169] start codon is not identified yet" /codon_start=1 /transl_table=11 /product="UNKNOWN PROTEIN FROM 2D-PAGE (SPOT LM6) (FRAGMENT)." . /protein_id="BAAL6532.1" /db_xref="GI:1800055" /translation="GLNFVKDAGEKLDVATGQHKDDQAKYQEHLNKTKGIPDAK VNIQIADKATVTDGSLQSEAKELILVAVGNSIGIASYDD" complement(340. .498) /gene="ZK632.10" complement(340. .498) /gene="ZK632.10" /note="similar to [SwissProt Accession Number P34655]" /codon_start=1 /transl_table=11 /protein_id="BAAL6533.1" /db_xref="GI:1800056" /translation="MGFWRIVTITLPLGLVLLKGGKGFNIIILLTLLGIYIPGLIH AFWQTRD" 681. .980 /gene="HLYU" 681. .980 /gene="HLYU"	
FEATURES source	/note="similar to [SwissProt Accession Number P52695]" /codon_start=1 /transl_table=11 /product="TRANSCRIPTIONAL ACTIVATOR HLYU." /protein_id="BAAL6534.1" /db_xref="GI:1800057" /translation="MTSLAQASAEQAAALLKAMHPKRLILLCLMSGSGPGTSAGEL TRITGLSATSQHLARMDEGLDSQDRAQRILLYSIKNEAVNAIATLKNVYCP" complement(1561. .1965) /gene="hnsB" /note="synonym: stpA" complement(1561. .1965) /gene="hnsB" /note="similar to [SwissProt Accession Number P30017]" /codon_start=1 /transl_table=11 /product="DNA-BINDING PROTEIN STPA." /protein_id="BAAL6535.1" /db_xref="GI:1800058" /translation="MSVMSLOANNIPTLRAMAREFFSIDVLEMLEKFRVVTKERREE EQQELAEARQEKISTWLELWKADGINPELLGNSSAAAPRAGKRQPRFAYKFTDV NGETKTWTGGRTPKPIAQALEGKSLDDFLI" complement(3120. .3476) /gene="ygaC" complement(3120. .>3476) /gene="ygaC" /note="similar to [SwissProt Accession Number P36931] start codon is not identified yet" /codon_start=1 /transl_table=11 /protein_id="BAAL6536.1" /db_xref="GI:1800059" /translation="RGSNNYLRDPDEVARVLEKVGFTVDVVTQKAYVRRGENYVYNR EARMGRTALVIHPTLKERSTLAEPASDIKTDHYQQFPLYLAGERHEHYGPHGFS RVALERYINGLFGAS" 3604. .3945 /gene="yqiJ" 3604. .3945 /gene="yqiJ" /note="similar to [SwissProt Accession Number P42617]" /codon_start=1 /transl_table=11 /protein_id="BAAL6537.1" /db_xref="GI:1800060" /translation="MGDHMFNRNDVDGVQDIQNDVNLADSLSVLXSWGSDAK GEAPARSKAQAALLKETARMHGRTRVQQAARDAVGCADSFVRERPWCSCVGTAAAVGI FIGALLSMRYS" 4193. .4846 /gene="MG230" 4193. .4846 /gene="MG230" /note="similar to [SwissProt Accession Number P47472]" /codon_start=1 /transl_table=11 /protein_id="BAAL6538.1" /db_xref="GI:1800061" /translation="WRITITRNDVCVOCHATKRAMENRGDFEMINVDVPEAEALR AQQFRQPVVIAGLDSWGFPMINFLASATRGQCQSLVYFSSSENTQRFIERL GLPAVRIPLNERERIQVDEPYLITVPSYGGGTAGAVPRQVIRFLNDEHNRALLRGVI ASGNRNFGEAYGRAGDVIARKCGVPWLRFELMGTSQDIENVRKGVTEFWQRQPQNA" 4858. .6414 /gene="NRDE" 4858. .6414 /gene="NRDE" /note="similar to [SwissProt Accession Number Q08698]" /codon_start=1 /transl_table=11 /product="RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE 2 ALPHA CHAIN (EC 1.17.4.1) (RIBONUCLEOTIDE REDUCTASE 2)." /protein_id="BAAL6539.1" /db_xref="GI:1800062" /translation="MDYHALNAMLNYDSAGRIQFDKDRQAVDAFIATHVRPNSVTFSSQQRLNLWLVNEGYSVILNRYSRDFVITLFTHTAHTSGFRFQFLGAWKYFTSYTLK	

Qy 1 MetSerTyrGluValLeuLeuGlyLeuValGlyValAlaAsnTyrCysPheArg 20
 Db 84241 ATGAGCTATGAGGTTCTGCTGGTGTACTGGTTGGCGGCGAATTATTGCTTCGC 84300
 Qy 21 TyrLeuProLeuArgLeuArgValGlyAsnAlaArgProThrIysArgGlyAlaValGly 40
 Db 84301 TATTGGCGCTCGCGCTCGGTGGTAAATGCCGCCCAACCAACGTCGCGGTAGT 84360
 Qy 41 IleLeuLeuAspThrIleGlyIleAlaSerIleCysAlaLeuValValSerThrAla 60
 Db 84361 ATTTTGCTCGACACATTGGCATCGCTCGATATGGCTCTGCTGGTGTCTCTACCGCA 84420
 Qy 61 ProGluValMetHisAspThrArgPheValProThrLeuValGlyPheAlaValLeu 80
 Db 84421 CCAGAAGTGATGACGATACACGCGTTCGTGGCCACGCTGCTCGCGTACTG 84480
 Qy 81 GlyAlaSerPheTyrIysThrArgSerIleIleIleProThrLeuLeuSerAlaLeuAla 100
 Db 84481 GGTGCCAGTTTCTATAAACACGCGACGATTTATCATCCCAACACTGCTTAGTGGCTGCC 84540
 Qy 101 TyrGlyLeuAlaTrpLysValMetAlaIleIle 111
 Db 84541 TATGGGCTCGCTGGAAAGTGATGGCGATTATA 84573

RESULT 9
 WPCOMMENT

Sequence split into 55 fragments LOCUS BA000007 Accession BA000007

Fragment Name	Begin	End
BA000007_00	1	110000
BA000007_01	100001	210000
BA000007_02	200001	310000
BA000007_03	300001	410000
BA000007_04	400001	510000
BA000007_05	500001	610000
BA000007_06	600001	710000
BA000007_07	700001	810000
BA000007_08	800001	910000
BA000007_09	900001	1010000
BA000007_10	1000001	1110000
BA000007_11	1100001	1210000
BA000007_12	1200001	1310000
BA000007_13	1300001	1410000
BA000007_14	1400001	1510000
BA000007_15	1500001	1610000
BA000007_16	1600001	1710000
BA000007_17	1700001	1810000
BA000007_18	1800001	1910000
BA000007_19	1900001	2010000
BA000007_20	2000001	2110000
BA000007_21	2100001	2210000
BA000007_22	2200001	2310000
BA000007_23	2300001	2410000
BA000007_24	2400001	2510000
BA000007_25	2500001	2610000
BA000007_26	2600001	2710000
BA000007_27	2700001	2810000
BA000007_28	2800001	2910000
BA000007_29	2900001	3010000
BA000007_30	3000001	3110000
BA000007_31	3100001	3210000
BA000007_32	3200001	3310000
BA000007_33	3300001	3410000
BA000007_34	3400001	3510000
BA000007_35	3500001	3610000
BA000007_36	3600001	3710000
BA000007_37	3700001	3810000
BA000007_38	3800001	3910000
BA000007_39	3900001	4010000
BA000007_40	4000001	4110000
BA000007_41	4100001	4210000
BA000007_42	4200001	4310000

BA000007_43 4300001 4410000
 BA000007_44 4400001 4510000
 BA000007_45 4500001 4610000
 BA000007_46 4600001 4710000
 BA000007_47 4700001 4810000
 BA000007_48 4800001 4910000
 BA000007_49 4900001 5010000
 BA000007_50 5000001 5110000
 BA000007_51 5100001 5210000
 BA000007_52 5200001 5310000
 BA000007_53 5300001 5410000
 BA000007_54 5400001 5498450

Continuation (36 of 55) of BA000007 from base 3500001 (BA000007 Escherichia coli O157:1)

Alignment Scores:
 Pred. No.: 1-75e-47 Length: 110000
 Score: 548.00 Matches: 110
 Percent Similarity: 99.1% Conservative: 0
 Best Local Similarity: 99.1% Mismatches: 1
 Query Match: 99.3% Indels: 0
 DB: 1 Gaps: 0

US-10-073-293A-6 (1-111) x BA000007_35 (1-110000)

Qy 1 MetSerTyrGluValLeuLeuGlyLeuValGlyValAlaAsnTyrCysPheArg 20
 Db 34502 ATGAGCTATGAGGTTCTGCTGGTACTGGTTGGCGGCGAATTATTGCTTCGC 34561
 Qy 21 TyrLeuProLeuArgLeuArgValGlyAsnAlaArgProThrIysArgGlyAlaValGly 40
 Db 34562 TATTGGCGCTCGCGCTCGGTGGTAAATGCCGCCCAACCAACGTCGCGGTAGT 34621
 Qy 41 IleLeuLeuAspThrIleGlyIleAlaSerIleCysAlaLeuValValSerThrAla 60
 Db 34622 ATTTGCTGCACACCATTTGGCATGCGCTCATATGCGCTCTGCTGGTGTCTTACCGCA 34681
 Qy 61 ProGluValMetHisAspThrArgPheValProThrLeuValGlyPheAlaValLeu 80
 Db 34682 CCAGAAGTGATGACGATACACGCGTTCGTGGCCACGCTGTCGCGGTACTG 34741
 Qy 81 GlyAlaSerPheTyrIysThrArgSerIleIleIleProThrLeuLeuSerAlaLeuAla 100
 Db 34742 GGTGCCAGTTTCTATAAACACGACGATTTATCATCCCAACACTGCTTAGTGGCTGCC 34801
 Qy 101 TyrGlyLeuAlaTrpLysValMetAlaIleIle 111
 Db 34802 TATGGGCTCGCTGGAAAGTGATGGCGATTATA 34834

RESULT 10

AE016987

LOCUS Shigella flexneri 2a str. 2457T DNA linear BCT 22-APR-2003
 DEFINITION Shigella flexneri 2a str. 2457T section 10 of 16 of the complete genome.

ACCESSION AE016987 AE014073

VERSION AE016987.1 GI:30042162

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

290380 bp DNA linear BCT 22-APR-2003
 Shigella flexneri 2a str. 2457T section 10 of 16 of the complete genome.
 AE016987 AE014073
 AE016987.1 GI:30042162
 Shigella flexneri 2a str. 2457T
 Shigella flexneri 2a str. 2457T
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Shigella.
 1 (bases 1 to 290380)
 Fournier, G., Mayhew, G.F., Plunkett, G. III, Rose, D.J., Darling, A.,
 Mau, B., Perna, N.T., Payne, S.M., Runyen-Janecky, L.J., Zhou, S.,
 Schwartz, D.C. and Blattner, F.R.
 Complete Genome Sequence and Comparative Genomics of Shigella
 flexneri Serotype 2a Strain 2457T
 Infect. Immun. 71 (5), 2775-2786 (2003)
 12704152
 2 (bases 1 to 290380)
 Wei, J., Goldberg, M.B., Burland, V., Venkatesan, M.M., Deng, W.,
 Fournier, G., Mayhew, G.F., Plunkett, G. III, Rose, D.J., Darling, A.,
 Mau, B., Perna, N.T., Payne, S.M., Runyen-Janecky, L.J., Zhou, S.,

TITLE JOURNAL FEATURES source	Schwartz, D. C. and Blattner, F. R. Direct Submission Submitted (13-JUN-2002) Genetics Laboratory, University of Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA Location/Qualifiers 1. .290380 /organism="Shigella flexneri 2a str. 2457T" /mol_type="genomic DNA" /strain="2457T" /serotype="2a" /db_xref="taxon:198215" complement (82. .1209) /gene="gcpE" /locus_tag="S2733" complement (82. .1209) /gene="gcpE" /locus_tag="S2733" /note="residues 1 to 372 of 375 are 96.23 pct identical to residues 1 to 372 of 372 from Escherichia coli K-12 : B2515"		gene	complement (3838. .4269) /gene="ndk" /locus_tag="S2736" complement (3838. .4269) /gene="ndk" /locus_tag="S2736" /function="enzyme; Purine ribonucleotide biosynthesis" /note="residues 1 to 143 of 143 are 100.00 pct identical to residues 1 to 143 of 143 from Escherichia coli K-12 : B2518"	
	/transl_table=1 /product="nucleoside diphosphate kinase" /protein_id="AAP17891.1" /db_xref="GI:30042166" /translation="MAIERTFSIIKPNVAKNVIGNIFARFEAAGFKIVGTQKMLHLTV EQAGFYAHDGKPFDFGLVFMFTSGPIVSVLEGENAVORHRLDGLGATNPANALAGT LRADYADSLTENGTHGSDSVESAAREIAYFFGEVGEVCPTR" complement (4418. .6517) /gene="pbpC" /locus_tag="S2737" /note="truncated" /pseudo 6514. .7961 /note="insertion element" /insertion_seq="ISSf14" complement (6534. .7730) /locus_tag="S2738" complement (6534. .7730) /locus_tag="S2738" /function="IS, phage, Tn; Transposon-related functions" /note="residues 1 to 398 of 398 are 90.95 pct identical to residues 1 to 398 of 398 from GenPept : >gp AAK18438.1 AF348706.127 (AF348706) putative transposase for IS110 [Shigella flexneri]" /codon_start=1 /transl_table=1 /product="ISSf14 orf" /protein_id="AAP17892.1" /db_xref="GI:30042167" /translation="MTESSDYESVQVFIGVDVGKDTTHAVAINRSGKRLFDKALPNDE NKRLSDIHKHQGILLVVDPATIGALPVAVARSEGLVGLVLPGLAMRRRLADLHAG EAKTVRDAALIAEAARTLPHALFTLKLADQIAELSMCLCGFDDDLAQTQASNRIR GLLTQIHPAFERVGLPRHPAULDLQRIIPSPKLSGKLLAQOLCKLAPLGRK LAADIAQALAEQTVVPGTNAUVPRLALQLITLRKQKDEVALEVEQVRVLAHLPLP VLTSMPGVGVRTAARLLTEVACRAFASAAHLAAYAGLAPVTRRSSSIRGEHPRSRRGN KALKRALFLSAFALRDLPLSRAYYTRKMSQGRHNQALIALARRCCDVLFAAMRDGT YTPQGS" 7960. .8442 /gene="sseA" /locus_tag="S2740" /note="truncated" /pseudo complement (9261. .10037) /gene="sseB" /locus_tag="S2741" complement (9261. .10037) /gene="sseB" /locus_tag="S2741" /function="phenotype; Not classified" /note="residues 1 to 258 of 258 are 94.96 pct identical to residues 4 to 261 of 261 from Escherichia coli K-12 : B2522"			complement (9261. .10037) /gene="sseB" /locus_tag="S2741" complement (9261. .10037) /gene="sseB" /locus_tag="S2741" /function="phenotype; Not classified" /note="residues 1 to 258 of 258 are 94.96 pct identical to residues 4 to 261 of 261 from Escherichia coli K-12 : B2522"	
gene	complement (1236. .2249) /gene="yfgA" /locus_tag="S2734" complement (1236. .2249) /gene="yfgA" /locus_tag="S2734" /function="putative membrane; Not classified" /note="residues 1 to 337 of 337 are 80.71 pct identical to residues 1 to 337 of 337 from Escherichia coli K-12 : B2516"		CDS	complement (9261. .10037) /gene="sseB" /locus_tag="S2741" complement (9261. .10037) /gene="sseB" /locus_tag="S2741" /function="phenotype; Not classified" /note="residues 1 to 258 of 258 are 94.96 pct identical to residues 4 to 261 of 261 from Escherichia coli K-12 : B2522"	
	/transl_table=1 /product="putative membrane protein" /protein_id="AAP17889.1" /db_xref="GI:30042164" /translation="MNTATHDNEALTTGARLNAREQLGLSQQAVERLCLKVSTV RDIREDKAPADLASTFLRGYIRSVARLVHIPEELLPGLEKQAPLRAAKVAPMQSFSL GRRKRGDLMTTLVLFVYVIGLSGAWWQDHKAQOEIITWADQSSAELSNSEQ GQVPLNTSTTPTTSPASVDTTATNTQTFAVATAPAVDPQONAVVPQANV DTAATPVTAATPDGAAPLPTDQAGVTTTAAADPNALVMNFTDQCLEVTDATGKKLF SGMQRKDGNLNLGQAPYKLGIGAPAAVQIQYQKPDLSRFIRTNQVARTLNAEQS PAQ" complement (2534. .3688) /gene="yfgB" /locus_tag="S2735" complement (2534. .3688) /gene="yfgB" /locus_tag="S2735" /note="residues 1 to 384 of 384 are 99.47 pct identical to residues 1 to 384 of 384 from Escherichia coli K-12 : B2517"			complement (9261. .10037) /gene="sseB" /locus_tag="S2741" complement (9261. .10037) /gene="sseB" /locus_tag="S2741" /function="phenotype; Not classified" /note="residues 1 to 258 of 258 are 94.96 pct identical to residues 4 to 261 of 261 from Escherichia coli K-12 : B2522"	
gene	complement (1236. .2249) /gene="yfgA" /locus_tag="S2734" complement (1236. .2249) /gene="yfgA" /locus_tag="S2734" /function="putative membrane; Not classified" /note="residues 1 to 337 of 337 are 80.71 pct identical to residues 1 to 337 of 337 from Escherichia coli K-12 : B2516"		CDS	complement (9261. .10037) /gene="sseB" /locus_tag="S2741" complement (9261. .10037) /gene="sseB" /locus_tag="S2741" /function="phenotype; Not classified" /note="residues 1 to 258 of 258 are 94.96 pct identical to residues 4 to 261 of 261 from Escherichia coli K-12 : B2522"	
	/transl_table=1 /product="hypothetical protein" /protein_id="AAP17890.1" /db_xref="GI:30042165" /translation="MSEQLVTPENVTTDKGINLLDLNRQMRBEFFKDLGKTFPRADQ VKRMWYHCCDNFDEMTDINKVLGKLEVAEIRAPVEVVEQRSSDGTGIKWAIVGDDQ RVETVYIPEDDRATLCVSSQVGCALCEKFCSTAQQGNRNLRVSEIIGQVWRAAKIG AAKVTGQRPITNVVMNGEPLLNINVPVAMEIMLDDFGFLSKRRVLTSTGVPVA LDLGLDMIDVALAISLHAPNDEIRDEIVPINKKNIETFLAAVRRYLEKSNANQGRVT L			complement (9261. .10037) /gene="sseB" /locus_tag="S2741" complement (9261. .10037) /gene="sseB" /locus_tag="S2741" /function="phenotype; Not classified" /note="residues 1 to 258 of 258 are 94.96 pct identical to residues 4 to 261 of 261 from Escherichia coli K-12 : B2522"	

LPTIKVPKRAPICSIKENEAQNLLIGIEADGDIEIIQATGSVATDTLPFGDEPID
ICQVKGEGKISHFITTEHAPFYERRWGGFLRDFKQNR11
10137. .10170
/note="REP (repetitive extragenic palindromic) element,
contains 1 REP sequence"
complement (10179. .11462)
/gene="pepb"
/locus_tag="S2742"
complement (10179. .11462)
/gene="pepb"
/locus_tag="S2742"
/function="putative enzyme; Not classified"
/note="residues 1 to 427 of 427 are 92.50 pct identical to
residues 30 to 456 from Escherichia coli K-12 ;
B2523"
/codon_start=1
/transl_table=11
/product="putative peptidase"
/protein_id="AAP17894.1"
/db_xref="GI:30042169"
ORARKIDGKIHQVLSGEGWADRCWAFQGYKPGKTRKVEWPLDDAQOELDN
RLMIIDWRDTINAPBELGPSOLAQRAVDLISNVAGDRVYRTTKGEDLRQGYMGL
HTVGRGERSPVLLADYNPTGDKAEAPVACLVCKGKITPDSGGYSIKQTAPMSMKSD
MGNAATYGALAFALMRGLNKRKVKFLCCADNLISGNAPFLGDDIITTRNGKKVEVMNT
DAEGLVLDGLIDASAQKPEMIIIDAATLTGAATGALNDYHALFSPDDALAGRLLAS
ASQNEPFWLPLAEFHRSQLPSNFASLNNITGSAAYPAGASATAGFLSHFVENYQQGW
LHIDCSATYRKAPVEQWSAGATGLGVRTIANLLTA"
complement (11521. .11721)
/gene="yfhJ"
/locus_tag="S2743"

Alignment Scores:
Pred. No.: 5.06e-47 Length: 290380
Score: 548.00 Matches: 110
Percent Similarity: 99.1% Conservative: 1
Best Local Similarity: 99.1% Mismatches: 1
Query Match: 99.3% Indels: 0
DB: 1 Gaps: 0

US-10-073-293A-6 (1-111) x AE016987 (1-290380)

Qy 1 MetSerTyrGluValLeuLeuLeuGlyLeuValGlyValAlaAenTyrCysPheArg 20
Db 155240 ATGAGCTATGAGGTTCTGCTGCTGGGTACTGTTGGCGCGCGCAATTATTGCTTCGCG 155299

Qy 21 TyrLeuProLeuArgLeuArgValGlyAsnAlaArgProThrLysArgGlyAlaValGly 40
Db 155300 TATTGGCGCTCGCGCTGGGTGGGTAAATGCCCGCCCAACCAACGTCGCGGTAGGT 155359

Qy 41 IleLeuLeuAspThrIleGlyIleAlaSerIleCysAlaLeuLeuValSerThrAla 60
Db 155360 ATTTGCTCGACACCATGTCATCGCTCGATATGCGCTCTGCTGGTGTCTCTACCGCA 155419

Qy 61 ProGluValMetHisAspThrArgArgPheValProThrLeuValGlyPheAlaValLeu 80
Db 155420 CCAGAACTGATGACAGTACACGCGCTTCTGTCGCCACGCTGCTCGCGTACTG 155479

Qy 81 GlyAlaSerPheThrLysThrArgSerIleIleIleProThrLeuSerAlaLeuAla 100
Db 155480 GGTGCCAGTTTCTATAAAACGACGACATTATATCATCCCAACACTGCTTAGTGGCTGGCC 155539

Qy 101 TyrGlyLeuAlaTrpLysValMetAlaIleIle 111
Db 155540 TATGGGCTCGCTGGAAAGTGATGGCGATTATA 155572

RESULT 11
AE016765
LOCUS
DEFINITION Escherichia coli CFT073 section 11 of 18 of the complete genome.
ACCESSION AE016765 AE014075
VERSION AE016765.1 GI:26109395
KEYWORDS

Escherichia coli CFT073
Escherichia coli CFT073
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
1 (bases 1 to 305325)
Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,
Raeko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,
Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,
Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,P.R.
Extensive Mosaic Structure Revealed by the Complete Genome Sequence
of Uropathogenic Escherichia coli
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 17020-17024 (2002)
12471157
2 (bases 1 to 305325)
Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,
Raeko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,
Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,
Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,P.R.
Direct Submission
Submitted (20-JUN-2002) Genetics Laboratory, University of
Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
source
1. 305325
/organism="Escherichia coli CFT073"
/mol_type="genomic DNA"
/strain="CFT073"
/db_xref="taxon:199310"
complement (145. .438)
/locus_tag="c3147"
complement (145. .438)
/locus_tag="c3147"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAN81599.1"
/db_xref="GI:26109396"
/translation="MDITPFLHALCAVAQLVLFTGNWAYGAIAGCTFFIAREHTQ
AEYRWIEMFGHGKRIINPMWGGFPRVWDVASLMDFAVPVVCALLVLLVNRG"
complement (481. .1521)
/locus_tag="c3148"
complement (481. .1521)
/locus_tag="c3148"
/notes="Residues 110 to 233 of 346 are 31.25 pct identical
to residues 486 to 611 of 669 from GenPept.129 ;
>gb|AAD03286.1| (AF063097) gph [Enterobacteria phage P2]"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAN81600.1"
/db_xref="GI:26109397"
/translation="MATSTVIPDDIKTLKSDVSKLNDQGSYATKLVDYSKDEIVGDM
SASWYQVLPSTGALFGRKLRSHTRTAGVEDAYCELYLKWIDSPGNAMARLNDNG
TNICWDFTLNYGVTMTIPGDSGYLKGKGNCLMSYSKSGSNALIIPDYITDTIQIKYANHG
STMLTNQGTAGVTAQLWGNSSRPVYEVGVDGAYMFYAQKNTDNTMYLSYNGAC
HATAFNQHSRDRLKNIQVIDNATIRKMGYTYTLKENGMPYAGVIAQEALEAIPR
VVSAMKYQDQSGASGSEGEERYTYDYSVTGLLVQVARESDDRITALEEENALRQ
RLSAIEAALASK"
complement (1531. .1812)
/locus_tag="c3149"
complement (1531. .1812)
/locus_tag="c3149"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAN81601.1"
/db_xref="GI:26109398"
/translation="MKDLTKFADRADFSAFMESIGYDDDSMQGGLIDIVIGNVYKE
TGBLTEDGPGVCVKEDGYFVNVRRIINDAKKSSIFDKYAIIVVEHLRGM"
complement (1809. .4238)
/locus_tag="c3150"
complement (1809. .4238)
/locus_tag="c3150"
/notes="Residues 118 to 438 of 809 are 30.28 pct identical

to residues 291 to 638 of 973 from GenPept.129 :
>gb|AAC56007.1|AE005333.4 (AE005333) putative membrane
protein of prophage CP-933X [Escherichia coli O157:H7
EDL9331]
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AA081602.1"
/db_xref="GI:26109399"
/translation="MPEDIVCGVDMAAVKISGVLKOGAGKPIQNCTIQLKAKRNSTV
LVNTVASDENDEAGRYSMDEVGYQSYVILLVEGPPSHAGTIYVEGPGTLDNDFG
AMTDEDDVMPALRFEEMBEAARNAEAASQSAAKKSETAAASKNAAKTSETNAA
NSAQAAATSTASNSATAAKSETNAKNGSETAAKTSETNAKSQTAAKTSETNAKAS
ETAKNSQVAAASQSAASGATSATGSAATAAASQAAKTSETNAKSQTAAKTSET
NAKASATAAKNSQDAQAASGASAAASASASASASANSQAAKTSETNAKASETHA
ANASAKAATAAKASAEADAREYASQAAPYKQVLPLEPDMVIFPNDSDLMITGSP
SYKIVGDIEMTWPBGDKVKFRKRASTATYINKSGVFSVAKIDEPREFEKGLLIEGQR
TNYVKGNDIAEWTSTNIDKNTNGVDFGFGFSAKMRTKDNMTQSSALSHTCSASR
GIDVSGNKYCTVSCRKADPLRCRLRFKYGDSVTFILGDAVLTETGTLIEKTGSA
ANRTAATKDPVTGWIFYEATIEAVSGETLIGAMIOYAPKKGVTGAGDIYILATPQ
FENGCCASSFVITTTIPATRSSDMVTIPTENNINRRELTCLVEVNRWGDIPNVAER
IFDSGVPPIESITAFNTTEKYYQLYMQTYKASTSTYSSVPAGRTDVRKFTGGFN
IYSDGTRVSVNGBATKMTKTEWTGVTTRTFIRIGGQATSGTRHLFGLHRLRLWHKE
LTDQMGESIK"
3403..3843
/locus_tag="c3151"
3403..3843
/locus_tag="c3151"
/locus_tag="c3151"
/locus_tag="c3151"
/note="Residues 15 to 146 of 146 are 45.32 pct identical
to residues 36 to 174 of 206 from SwissProt.40 :
>sp|P03774|Y206_LAMB_HYPOTHETICAL PROTEIN ORF206"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AA081603.1"
/db_xref="GI:26109400"
/translation="MAFVSLVFAAFLVDFALVSLVFAAFWLLAAVAADPVAEVAEPAA
ALSIWAATWLPFAAVALAFVSLVFAAFLVDFALVSLVFAAFLVDFALVDFALV
AEFSDAACEVAACALLAAAFVSEVFAAFFDDAAVSDFFFAAAL"
complement(4222..4932)
/insertion_seq="IS1541A-like"
complement(4329..4838)
/locus_tag="c3152"
complement(4329..4838)
/locus_tag="c3152"
/note="Residues 18 to 169 of 169 are 95.39 pct identical
to residues 1 to 152 of 152 from SwissProt.40 :
>sp|Q57334|T200_SALTY Transposase for insertion sequence
element IS200"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AA081604.1"
/db_xref="GI:26109401"
/translation="MRSNCKCOTRNOKGVPMGNEKSLAHTRWNCYHIVPAPKYRQ
VFYREKRAAGSIIRKLCEWKSRIIEACCAADHMLVEIPPKMSVSGPMYLGKGS
SLMLYEQFGDILKYIRNRFWCGRGYVDTVGKNTAKIQDYIKHQLSEEDKMGESUISY
PGSPFTGRK"
complement(4984..5583)
/locus_tag="c3153"
complement(4984..5583)
/locus_tag="c3153"
/function="putative membrane; Other or Phage or Prophage
Related"
/note="Escherichia coli O157:H7 ortholog: z1917"
/codon_start=1
/transl_table=11
/product="putative outer membrane protein of prophage"
/protein_id="AA081605.1"
/db_xref="GI:26109402"
/translation="MRKLCAVILSAGVVLVNLVAAGTPASAAEHQSTLSAGYLQTHDMPG
SDDLKGINVKYRFTDTLGLVTSFSYANAKDEQKTHYSPTRMHEDSVNRNRFWMAG
PSVRVNEWFSAYAMAGWAYSRVTSFGDYLRVTDNKGKTHDVLVTGDDGRHSNTSLAW

gene
complement(5651..9130)
/locus_tag="c3154"
complement(5651..9130)
/locus_tag="c3154"
/function="putative
or Prophage Related"
/note="Escherichia coli O157:H7 ortholog: z2145"
/codon_start=1
/transl_table=11
/product="Putative tail component of prophage"
/protein_id="AA081606.1"
/db_xref="GI:26109403"
/translation="MGKSSKGHTPREAKDNLKSSOMLSVIDAISGEVPEGVVDGLKS
VLLNSPVLDSSEGNWISGVTVFRAGEQETPEGSESSGSETVLCTEYKYDPIIR
TITNSIDRURFTGVOALVETTSKGRNLSSEVRLLVQIRNGWGWTEKDIITKGTIT
VLSVSVVNDLPPRPNIRMRMTDSTTDQONKTLWSYSYIIIDVKQCPYNALV
GVQVDSVQSQSVSRNYHLRILQVPSNYPQTRQYISWDGTFPAPYNNMAWCL
WDLMTHTYRGMKRLGAADVQWALYVIGQVCDQSPDGFQGTPEPTICNAMLTTQRK
AWDLSDFCSAMRCPVWNGSOTLTFCQDRPSDKVMTYRNVNVPMDGAPFRYFSAL
KDRHNAVENWIDPDGWEATLVEDTQAIARYGNVTKMDAPGCTSRGQAHRLGLW
LIKTELLETVTDFSVGAEGRHVPGDIIIEICDDYAGISTGSRVLAVNSQGRKLTLID
REITLPSGGTTLISLVDSGNPVSQVSTVDYKVSVRVDPGVAQSGVTGTPPAVQHL
RQLRFRCVSIRENDDGTALTAQVHVEKEALVDNGAHFQSGSTVGTTPPAVQHL
TAEVTADSGEYQVLARWDTPKVVGVSFMRLLTVAADDGSERLVSTARTTETTYRFRQ
LALGSVELTVRAANAMGQQGDPASVSFRIAPAAPSRIELTPGVFOITATPHLAVYDP
TVQPEFWFSEKRIADIRQVETTYALTYWIAASINIKPGHDYFYIERSVNTVGS
APEAVGQSPDSDASGYLDFPFKEIGKSHLAQELWQIDNGQLAPDLAERTSITDVEN
EITQTVNKKLEDQSAAIQIQKVQVDTNNLNSMVAVKLQOMQDGRILYIAGIGAIEN
TPDQMSQVLLAADRIAMVNPANGTKPMFVGQGDQIFMNEVFLKYLITAPTITSGGNP
PASFLLPDRLTAKNADISGNVANSGTLNANNVINENCRVGLKLSAQIIEGLDKVTG
KAFPROSRAPERWPSGTITVRIYDDQPFDRQIVIPAVAFSGAKHERHTDIYSSCRLL
VRNGAEIYNRTALDNTLIYSGVIDMPAGHGMTLEFSVSAWLNNWYPTASISDLLV
VVMKATAGISIS"
complement(9191..9823)
/locus_tag="c3155"
complement(9191..9823)
/locus_tag="c3155"
/function="putative structure; Structural component Phage
or Prophage Related"
/note="Escherichia coli O157:H7 ortholog: z0979"
/codon_start=1
/transl_table=11
/product="Putative tail component of prophage"
/protein_id="AA081607.1"
/db_xref="GI:26109404"
/translation="MARICLYDLQRFGRIDIRKVTGABAIRALATQLPAPROKLS
GWYQVRIAGRDVSTSGTLAQHETLPDGAIVHIVPRVAGAKSGGVFOIVLGAAGIAGS
FFTAGATLAAWGAAGAGCGMTGILFSLGASNVLGGVAQMLAPKARTPRTOTTDNGKQK

Alignment Scores:
Pred. NO.: 5.34e-47 Length: 305325
Score: 548.00 Matches: 110
Percent Similarity: 99.1% Conservative: 0
Best Local Similarity: 99.1% Mismatches: 1
Query Match: 99.3% Indels: 0
DB: 1 Gaps: 0

US-10-073-293A-6 (1-111) x AE016765 (1-305325)

Qy 1 MetSerTyrGluValLeuLeuLeuGlyLeuValGlyValAlaAAsnTyrCysPheArg 20
Db 68932 ATGAGCTATGAGGTTCTGCTGCTTGGGTACTGTTGGCGGCGGAATTATTGCTTCGC 68931
Qy 21 TyrLeuProLeuArgLeuArgValGlyAsnAlaArgProThrLysArgGlyAlaValGly 40
Db 68992 TATCTCGCGTGCCTGCGTGGGTAAATGCCCGCCCAACCAACGTCGCGCTAGGT 68951
Qy 41 IleLeuLeuAspThrIleGlyIleAlaSerIleCysAlaLeuLeuValSerThrAla 60
Db 68952 ATTTTGTCCGACACATTTGGCATCGCTCGATATGCGCTCTGCTGCTGTCTTACCCCA 69011
Qy 61 ProGluValMetHisAspThrArgPheValProThrLeuValGlyPheAlaValLeu 80

Db 69012 CCAGAGGTGATGACGACGCGTTTCGTGGCCACACTGGTGGCGTTCGCGGTACTG 69071
Qy 81 GlyAlaSerPheTyrLysThrArgSerIleIleIleProThrLeuLeuSerAlaLeuAla 100
Db 69072 GGTGCCAGTTTCTATATAAACACGACGACGATTATCATCCCAACACTGCTTAGTGGCGTGGCC 69131
Qy 101 TyrGlyLeuAlaTrpLysValMetAlaIleIle 111
Db 69132 TATGGGCTCGCCTGGAAAGTGATGGCGATTATA 69164

RESULT 12

AE005174_35

WPCOMMENT

Sequence split into 56 fragments LOCUS AE005174 Accession AE005174

Fragment Name	Begin	End
AE005174_00	1	110000
AE005174_01	100001	210000
AE005174_02	200001	310000
AE005174_03	300001	410000
AE005174_04	400001	510000
AE005174_05	500001	610000
AE005174_06	600001	710000
AE005174_07	700001	810000
AE005174_08	800001	910000
AE005174_09	900001	1010000
AE005174_10	1000001	1110000
AE005174_11	1100001	1210000
AE005174_12	1200001	1310000
AE005174_13	1300001	1410000
AE005174_14	1400001	1510000
AE005174_15	1500001	1610000
AE005174_16	1600001	1710000
AE005174_17	1700001	1810000
AE005174_18	1800001	1910000
AE005174_19	1900001	2010000
AE005174_20	2000001	2110000
AE005174_21	2100001	2210000
AE005174_22	2200001	2310000
AE005174_23	2300001	2410000
AE005174_24	2400001	2510000
AE005174_25	2500001	2610000
AE005174_26	2600001	2710000
AE005174_27	2700001	2810000
AE005174_28	2800001	2910000
AE005174_29	2900001	3010000
AE005174_30	3000001	3110000
AE005174_31	3100001	3210000
AE005174_32	3200001	3310000
AE005174_33	3300001	3410000
AE005174_34	3400001	3510000
AE005174_35	3500001	3610000
AE005174_36	3600001	3710000
AE005174_37	3700001	3810000
AE005174_38	3800001	3910000
AE005174_39	3900001	4010000
AE005174_40	4000001	4110000
AE005174_41	4100001	4210000
AE005174_42	4200001	4310000
AE005174_43	4300001	4410000
AE005174_44	4400001	4510000
AE005174_45	4500001	4610000
AE005174_46	4600001	4710000
AE005174_47	4700001	4810000
AE005174_48	4800001	4910000
AE005174_49	4900001	5010000
AE005174_50	5000001	5110000
AE005174_51	5100001	5210000
AE005174_52	5200001	5310000
AE005174_53	5300001	5410000
AE005174_54	5400001	5510000
AE005174_55	5500001	5528445

Continuation (36 of 56) of AE005174 from base 3500001 (AE005174 Escherichia coli O157:H7)

Alignment Scores:

Pred. No.: 4.74e-47 Length: 110000
Score: 544.00 Matches: 109
Percent Similarity: 98.2% Conservative: 0
Best Local Similarity: 98.2% Mismatches: 2
Query Match: 98.6% Indels: 0
DB: 1 Gaps: 0

US-10-073-293A-6 (1-111) x AE005174_35 (1-110000)

Qy 1 MetSerTyrGluValLeuLeuGlyLeuValGlyValAlaAenTyrCysPheArg 20
Db 101792 ATGAGCTATGAGTTCTGCTGTGGTTACTGGTTGGCGCGGAATATTTGCTTCGCG 101851
Qy 21 TyrLeuProLeuArgLeuArgValGlyAsnAlaArgProThrLysArgGlyValAlaValGly 40
Db 101852 TATTTGGCGCTGGCGCTGCGTGTGGGTAAATGCCGCCCAACCAACGTTGGCGGTAGGT 101911
Qy 41 IleLeuLeuAspThrIleGlyIleAlaSerIleCysAlaLeuLeuValSerThrAla 60
Db 101912 ATTTTGCTCGACACCAATTGGCATCGCTCGATATGCGTCTGCTGGTTGTCTTACCGCA 101971
Qy 61 ProGluValMetHisAspThrArgArgPheValProThrLeuValGlyPheAlaValLeu 80
Db 101972 CCAGAGTGATGCACGATACACGCCGTTTCGTGCCACGCTGGTTCGCGGTACTG 102031
Qy 81 GlyAlaSerPheTyrLysThrArgSerIleIleIleProThrLeuLeuSerAlaLeuAla 100
Db 102032 GTTGCCAGTTTCTATATAAACACGACGATATATATCCCAACACTGCTTARTGCCGTGGCC 102091
Qy 101 TyrGlyLeuAlaTrpLysValMetAlaIleIle 111
Db 102092 TATGGGCTCGCCTGGAAAGTGATGGCGATTATA 102124

RESULT 13

AE005174_36

WPCOMMENT

Sequence split into 56 fragments LOCUS AE005174 Accession AE005174

Fragment Name	Begin	End
AE005174_00	1	110000
AE005174_01	100001	210000
AE005174_02	200001	310000
AE005174_03	300001	410000
AE005174_04	400001	510000
AE005174_05	500001	610000
AE005174_06	600001	710000
AE005174_07	700001	810000
AE005174_08	800001	910000
AE005174_09	900001	1010000
AE005174_10	1000001	1110000
AE005174_11	1100001	1210000
AE005174_12	1200001	1310000
AE005174_13	1300001	1410000
AE005174_14	1400001	1510000
AE005174_15	1500001	1610000
AE005174_16	1600001	1710000
AE005174_17	1700001	1810000
AE005174_18	1800001	1910000
AE005174_19	1900001	2010000
AE005174_20	2000001	2110000
AE005174_21	2100001	2210000
AE005174_22	2200001	2310000
AE005174_23	2300001	2410000
AE005174_24	2400001	2510000
AE005174_25	2500001	2610000
AE005174_26	2600001	2710000
AE005174_27	2700001	2810000
AE005174_28	2800001	2910000
AE005174_29	2900001	3010000
AE005174_30	3000001	3110000
AE005174_31	3100001	3210000
AE005174_32	3200001	3310000

AE005174_33 3300001 3410000
AE005174_34 3400001 3510000
AE005174_35 3500001 3610000
AE005174_36 3600001 3710000
AE005174_37 3700001 3810000
AE005174_38 3800001 3910000
AE005174_39 3900001 4010000
AE005174_40 4000001 4110000
AE005174_41 4100001 4210000
AE005174_42 4200001 4310000
AE005174_43 4300001 4410000
AE005174_44 4400001 4510000
AE005174_45 4500001 4610000
AE005174_46 4600001 4710000
AE005174_47 4700001 4810000
AE005174_48 4800001 4910000
AE005174_49 4900001 5010000
AE005174_50 5000001 5110000
AE005174_51 5100001 5210000
AE005174_52 5200001 5310000
AE005174_53 5300001 5410000
AE005174_54 5400001 5510000
AE005174_55 5500001 5528445
Continuation (37 of 56) of AE005174 from base 3600001 (AE005174 Escherichia coli O157:H7)

Alignment Scores:
Pred. No.: 4,74e-47 Length: 110000
Score: 544.00 Matches: 109
Percent Similarity: 98.2% Conservative: 0
Best Local Similarity: 98.2% Mismatches: 2
Query Match: 98.6% Indels: 0
DB: 1 Gaps: 0

US-10-073-293A-6 (1-111) x AE005174_36 (1-110000)

Qy 1 MetSerTyrGluValLeuLeuLeuGlyLeuValGlyValAlaAanTyrCysPheArg 20
Db 1792 ATGAGCTATGAGTTCCTGCTGGGTACTGTTGGCGGGCGGAATTATGCTTCGCG 1851

Qy 21 TyrLeuProLeuArgLeuArgValGlyAsnAlaArgProThrLysArgGlyAlaValGly 40
Db 1852 TATTGGCGCTGGCGCTCGGTGGTAAATGCCCGCCCAACCAAGCTGGCGGTAGGT 1911

Qy 41 IleLeuLeuAspThrIleGlyIleAlaSerIleCysAlaLeuValValSerThrAla 60
Db 1912 ATTTTGTGCGACACCATTTGGCATCGCTCGATATGCGCTCTGCTGGTGTCTCTACCGCA 1971

Qy 61 ProGluValMetHisAspThrArgArgPheValProThrLeuValGlyPheAlaValLeu 80
Db 1972 CCAGAGTGTGACGATACAGCGCGCTTGTGCGCCACGCTGGTTCGGGTACTG 2031

Qy 81 GlyAlaSerPheTyrLysThrArgSerIleIleIleProThrLeuLeuSerAlaLeuAla 100
Db 2032 GGTGCCAGTTTCTATAAAACACGACGAGCATTTATCATCCCAACTGCTTARTGGCGTGGCC 2091

Qy 101 TyrGlyLeuAlaTrrLysValMetAlaIleIle 111
Db 2092 TATGGCTCGCTGGAAAGTATGATCGCATATA 2124

RESULT 14
AR386318
LOCUS AR386318 372 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 3047 from patent US 6610836.
ACCESSION AR386318
VERSION AR386318.1 GI:40096052
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 372)
AUTHORS Breton,G.L. and Osborne,M.
TITLE Nucleic acid amino acid sequences relating to Klebsiella pneumoniae for diagnostics and therapeutics

JOURNAL Patent: US 6610836-A 3047 26-AUG-2003;
Genome Therapeutics Corporation; Waltham, MA
FEATURES
source Location/Qualifiers
1..372
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 1.7e-34 Length: 372
Score: 403.00 Matches: 77
Percent Similarity: 84.1% Conservative: 13
Best Local Similarity: 72.0% Mismatches: 17
Query Match: 73.0% Indels: 0
DB: 6 Gaps: 0

US-10-073-293A-6 (1-111) x AR386318 (1-372)

Qy 1 MetSerTyrGluValLeuLeuLeuGlyLeuValGlyValAlaAanTyrCysPheArg 20
Db 37 ATGACAGTGAAGTCCCTTCGTCGGGTGATAGTCGGCGGCTCAACTATCTCTTTCCG 96

Qy 21 TyrLeuProLeuArgLeuArgValGlyAsnAlaArgProThrLysArgGlyAlaValGly 40
Db 97 TATCTTCTCTGCTGCTTTGGCGCGGCACGCTGCTCTGCCCGCGCGGCTCAGC 156

Qy 41 IleLeuLeuAspThrIleGlyIleAlaSerIleCysAlaLeuValValSerThrAla 60
Db 157 GTGCTGCTGGATACTATTGGCATCGCTCGCATTTGCCGCTGCTGGTGGTCTCCAGCGTG 216

Qy 61 ProGluValMetHisAspThrArgArgPheValProThrLeuValGlyPheAlaValLeu 80
Db 217 CCGAGATCTTCGCGCATGCCCATCGCTGGCGCGGACGCTGACCGGCTTCTGGTGTG 276

Qy 81 GlyAlaSerPheTyrLysThrArgSerIleIleIleProThrLeuLeuSerAlaLeuAla 100
Db 277 GGGCGCGCTTCTGGAACCCGCGACGATCATTTGCCGACCTGCTGCGCGCTTGGCC 336

Qy 101 TyrGlyLeuAlaTrrLysVal 107
Db 337 TATGCTGTCATGGAATC 357

RESULT 15
ECH410307 3619 bp DNA linear BCT 15-APR-2005
LOCUS Erwinia chrysanthemi ygaz gene (partial), ygah gene, emrR gene,
DEFINITION emrA gene and emrB gene (partial).
ACCESSION AJ410307
VERSION AJ410307.1 GI:14970541
KEYWORDS emrA gene; EmrA protein; emrB gene; EmrB protein; emrR gene; EmrR protein; ygah gene; ygah protein; ygaz gene; ygaz protein.
SOURCE Erwinia chrysanthemi (Pectobacterium chrysanthemi)
ORGANISM Erwinia chrysanthemi
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Pectobacterium.
REFERENCE 1
AUTHORS Reverchon,S.
TITLE Characterization of the Erwinia chrysanthemi multidrug resistance pump EmrAB
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3619)
AUTHORS Reverchon,S.
TITLE Direct Submission
JOURNAL Submitted (12-JUL-2001) Reverchon S., Unite de Microbiologie et Genetique, Institut des Sciences Appliquees de Lyon, Batiment Louis Pasteur, 11 Avenue Jean Capelle, F69621 Villeurbanne cedex, FRANCE
FEATURES
source Location/Qualifiers
1..3619
/organism="Erwinia chrysanthemi"
/mol_type="genomic DNA"
/strain="3937"
/db_xref="taxon:556"
<1..746
gene

CDS

```

/gene="ygaz"
<1. .746
/gene="ygaz"
/function="unknown"
/codon_start=3
/transl_table=11
/product="ygaz protein"
/protein_id="CAC44347.1"
/db_xref="GI:14970542"
/db_xref="InterPro:IPR011606"
/db_xref="UniProt/TREMBL:Q93KB4"
/translations="NSVSSQADARTTADASFGSLFDSLPIVIGYVPVAFAGLNAY
KLGTPPEALFLSCIIYAGASQFVITALLSAGASIVWAALTYWAMDVRHVLVIGPSLR
RIGWRPTKTAWAFGLTDEVFAATLARLDRNRWSEPMGLGVALSWSVAGTV
IGAVRNGPLEGYPVPAALAFMLPALFLSLLASFRRSQSLVVAALGACGLGLVVS
SIPAAIIGIGGGCLASLYQPTSVKEETR"
743. .1084

```

gene

CDS

```

/gene="ygaH"
743. .1084
/gene="ygaH"
/function="unknown"
/codon_start=1
/transl_table=11
/product="ygaH protein"
/protein_id="CAC44348.1"
/db_xref="GI:14970543"
/db_xref="UniProt/TREMBL:Q93KB3"
/translations="MNTSVLLIGLVGVNPLFRYLPLRGLAGRTSGNLRGRRTALLL
DSIGIASICALLVSTLPDIMQHTKMLPLTAGFALLTLCFYKTRSVLSTLGCALCY
GIVFKNQPTWA"
1259. .1771

```

gene

CDS

```

/gene="emrR"
1259. .1771
/gene="emrR"
/function="negative regulator"
/codon_start=1
/transl_table=11
/product="EmrR protein"
/protein_id="CAC44349.1"
/db_xref="GI:14970544"
/db_xref="GOA:Q93KB2"
/db_xref="InterPro:IPR000835"
/db_xref="InterPro:IPR011991"
/db_xref="UniProt/TREMBL:Q93KB2"
/translations="MSSSFAPIEDMLRMASRRPDPYREVLLRLFLHMOTKILSHR
NRMLKQDINETFLMALLTLESQENYCIQPSLSAALGSWRTNATRIADELEKRGWIE
RRSDSDRRLYLMTKEKGAFLDELLPPQHRSLNILCSALEDSESDQLEVLNRKLLQ
RLDEMDDQGI"
1946. .3121

```

gene

CDS

```

/gene="emrA"
1946. .3121
/gene="emrA"
/functions="membrane fusion protein"
/codon_start=1
/transl_table=11
/product="EmrA protein"
/protein_id="CAC44350.1"
/db_xref="GI:14970545"
/db_xref="GOA:Q93KB1"
/db_xref="InterPro:IPR005694"
/db_xref="InterPro:IPR006143"
/db_xref="UniProt/TREMBL:Q93KB1"
/translations="MSANVEKQTPPTMPSKSKSRKGLTLLMIVFFIGCVWFAYVY
LVLRHQETDADVAGNQIQIMSQVSGSVARVNVNDTFVYKGDVLVELDPTDAEQAF
ERSKTTLASVRVHQHMVNRQIQANIDLQQLADKAIKSLNRREVGLGRANAIGRED
LQHQDQVASARASLEAKQOQYATQALVLNTLSLEQOPIAQAAATDLRNALWKETH
IVSPVDGVSRVQVGRVTPSSALMAVVPAAFLWVDANFKETQLANRIRIGQAPETII
SDLYGSDVYKGVGLDMGTGSAFLSPQANATGNWIKVVERLPRVRIELDPKQLTDY
PLRIGLSTLVKVDNTDQKWLSDASRATPAYQSDALDLDLSPVNMQIDQIIGANAG"
3139. .>3619

```

gene

CDS

```

/gene="emrB"
3139. .>3619
/gene="emrB"

```

```

/function="translocase"
/codon_start=1
/transl_table=11
/product="EmrB protein"
/protein_id="CAC44351.1"
/db_xref="GI:14970546"
/db_xref="GOA:Q93KB0"
/db_xref="InterPro:IPR001411"
/db_xref="InterPro:IPR007114"
/db_xref="InterPro:IPR011701"
/db_xref="UniProt/TREMBL:Q93KB0"
/translations="WTRKLEGMTLMTIALSLATFMQVLQDSTIANVAIPTAGNLG
ANSQGTWITTSFVANALISIPITGLAKRFGEVRLFIWATVLTFTLTSWICGMSSTLE
MLTVSRMLQGLVAGPIIPLSQSLNNYPKRGIALALWSMTVVVAFIPGPILGWI"
ORIGIN
Alignment Scores:
Pred. No.: 9,82e-24 Length: 3619
Score: 313.50 Matches: 63
Percent Similarity: 72.9% Conservative: 15
Best Local Similarity: 58.9% Mismatches: 28
Query Match: 56.8% Indels: 1
DB: Gaps: 1
US-10-073-293A-6 (1-111) x ECH410307 (1-3619)
Qy 1 MetSerTyrGluValLeuLeuLeuGlyLeuValGlyValAlaLeuTyrCysPheArg 20
Db 743 ATGAATCTTCAGTATTATTGATTGTTTACTGTGGGAACAGTTAACTTCTCTGTTCCGA 802
Qy 21 TyrLeuProLeuArgLeuArgValGlyAsnAlaArgProThr---LysArgGlyAlaVal 39
Db 803 TACCTGCGCTGCGGTTGGCGCAGCGCGACCTCCGGTAACCTGCGCGCGGCGAGAAC 862
Qy 40 GlyIleLeuLeuAspThrIleGlyIleAlaSerIleCysAlaLeuLeuValValSerThr 59
Db 863 GCCCTGCTGCTGCACAGCATCGGCATTGCGTCGATTGTCGCTGCTGTGTATCCACC 922
Qy 60 AlaProGluValMethIleAspThrArgPheValProThrLeuValGlyPheAlaVal 79
Db 923 CTCGCCGNATATCAGCACGACGACGAAAGATGTTACCGACGCTGGCGGGGTTTCGCCCTG 982
Qy 80 LeuGlyAlaSerPheTyrLysThrArgSerIleIleIleProThrLeuLeuSerAlaLeu 99
Db 983 CTGACGCTTTGTTTTTACAAAACCCGACGATCGTACTATCAACGCTGCTGGCGCGCTG 1042
Qy 100 AlaTyrGlyLeuAlaTrpLys 106
Db 1043 TGCTACGGCATAGTATTATAA 1063

```

Search completed: February 16, 2006, 19:47:18
Job time : 2895.6 secs

This Page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 16, 2006, 17:00:29 ; Search time 333.935 Seconds
(without alignments)
2215.341 Million cell updates/sec

Title: US-10-073-293A-6
Perfect score: 552
Sequence: 1 MSYEVLLGLLVGVNYCFR'.....IPTLLSALAYGLAWKVAII 111

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Deiop 6.0 , Delext 7.0

Searched: 4996997 seqs, 3332346308 residues
Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/abs/ABSSWEB spool/US10073293/runat_15022006_120733_951/app_query.fasta.1
-DB=N_Geneseq -QFMT=fasta -SURFEX=p2n.rng -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abs07
-USPR=US10073293 @CGN 1.1 900 @runat_15022006_120733_951 -NCPU=6 -ICPU=3
-NO_WMAP -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_21.*

- 1: Geneseq1980s.*
- 2: Geneseq1990s.*
- 3: Geneseq2000s.*
- 4: Geneseq2001as.*
- 5: Geneseq2001bs.*
- 6: Geneseq2002as.*
- 7: Geneseq2002bs.*
- 8: Geneseq2003as.*
- 9: Geneseq2003bs.*
- 10: Geneseq2003cs.*
- 11: Geneseq2003ds.*
- 12: Geneseq2004as.*
- 13: Geneseq2004bs.*
- 14: Geneseq2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	552	100.0	336	6	ABV75657 E. coli L
2	552	100.0	336	14	ADZ39952 Adz39952 E.coli yg
3	552	100.0	336	14	ADZ36196 Adz36196 Escherich
4	423	76.6	314	4	AAK70628 Human imm

5	403	73.0	372	11	ACH97252
6	325	58.9	238	4	AAK57571
7	271.5	49.2	339	10	ACF69126
8	271.5	49.2	110000	10	ACF67367_20
9	271.5	49.2	243072	10	ACF65382
10	182.5	33.1	369	9	ADA31451
11	164	29.7	100	8	ACD80498
12	160	29.0	100	8	ACD80497
13	155	28.1	100	8	ACD80499
14	98	17.8	172	10	ACF79478
15	78	14.1	1593	8	ACA39729
16	78	14.1	1879	8	ACA05386
17	78	14.1	1879	6	ABA00331
18	77.5	14.0	1326	8	ACA38890
19	77.5	14.0	19988	4	AAF28523
20	77	13.9	324	12	ADK13752
21	77	13.9	330	4	AAH84429
22	77	13.9	333	3	AAH66035
23	77	13.9	333	4	AAH52620
24	77	13.9	333	4	AAH84506
25	77	13.9	333	5	AAH81328
26	77	13.9	333	8	ACA18608
27	77	13.9	333	8	ACA13378
28	77	13.9	924	13	AD848786
29	77	13.9	10944	4	AAH46250
30	76	13.8	330	8	ACA49494
31	76	13.8	333	4	AAH56235
32	76	13.8	333	8	ACA19680
33	76	13.8	333	8	ACA51662
34	76	13.8	405	8	ACA49545
35	76	13.8	1544	6	ABX97164
36	76	13.8	1544	12	ADN62044
37	76	13.8	1780	3	ACA32850
38	76	13.8	1950	8	ACA46614
39	76	13.8	1968	6	ABN93231
40	76	13.8	1968	13	ADS04188
41	76	13.8	4041	4	AAH54209
42	76	13.8	4342	4	AAH54625
43	75.5	13.7	999	9	ADB11933
44	75.5	13.7	999	9	ADB11925
45	75.5	13.7	999	9	ADB11921

ALIGNMENTS

RESULT 1	
ABV75657	ABV75657 standard; DNA; 336 BP.
ID	ABV75657
AC	ABV75657;
XX	
DT	29-JAN-2003 (first entry)
XX	
DE	E. coli L-amino acid producing gene #2.
XX	
KW	L-amino acid; gene; ss; E.coli.
XX	
OS	Escherichia coli.
XX	
PH	Key
FT	CDS
FT	1..336
XX	/*tag= a
PN	EP1239041-A2.
XX	
PD	11-SEP-2002.
XX	
PF	13-FEB-2002; 2002EP-00003335.
XX	
PR	13-FEB-2001; 2001RU-00103865.
PR	26-FEB-2001; 2001RU-00104998.
PR	26-FEB-2001; 2001RU-00104999.
PR	28-JUN-2001; 2001RU-00117632.

Qy 61 ProGluValMetHisAspThrArgPheValProThrLeuValGlyPheAlaValLeu 80
 Db 181 CCAGAAGTGATGCACGATACACGCGTTTCGTGCCACGCTGCTGCGGTACTG 240
 Qy 81 GlyAlaSerPheThrArgSerIleIleProThrLeuLeuSerAlaLeuAla 100
 Db 241 GGTGCCAGTTTCTATAAACACGACGATTCATCCACACACTGCTTAGTGGCGCTGCC 300
 Qy 101 TyrGlyLeuAlaTrpLysValMetAlaIleIle 111
 Db 301 TATGGCTCGCTGGAAGTGATGCCGATTATA 333

RESULT 3

ADZ36196

ID ADZ36196 standard; DNA; 336 BP.

XX AC ADZ36196;

XX 14-JUL-2005 (first entry)

XX Escherichia coli b2683 encoding DNA SEQ ID NO:5.

XX amino acid production; gene; ds.

XX Escherichia coli.

XX Key Location/Qualifiers

XX CDS 1..336

XX /*tag= a

XX /product= "b2683 protein"

XX EPI526181-A1.

XX 27-APR-2005.

XX 13-FEB-2002; 2004EP-00028877.

XX 13-FEB-2001; 2001RU-00103865.

XX 26-FEB-2001; 2001RU-00104998.

XX 26-FEB-2001; 2001RU-00104999.

XX 28-JUN-2001; 2001RU-00117632.

XX 28-JUN-2001; 2001RU-00117633.

XX 13-FEB-2002; 2002EP-00003335.

XX (AJIN) AJINOMOTO CO INC.

XX Tabolina EA, Rybak KV, Khourges EM, Voroshilova EB, Guseyatiner MM;

XX WPI; 2005-317157/33.

XX P-PSDB; ADZ36197.

XX New modified L-amino acid producing bacterium useful to enhance the

XX production of L-amino acid by enhancing the activities of proteins in a

XX cell of bacterium.

XX Disclosure; SEQ ID NO 5; 35pp; English.

XX The invention relates to an L-amino acid (A) producing bacterium (I)

XX (belonging to the genus Escherichia), which is modified to enhance the

XX production of (A) by enhancing the activities of proteins (G) or (H) in a

XX cell of (I). Also described is a method for producing (A) comprising

XX cultivating the bacterium in a culture medium and collecting the produced

XX and accumulated L-amino acid. The modified bacterium has the ability to

XX grow on a minimal medium containing L-amino acid or its analog in a

XX minimal concentration and ability to grow faster on a medium containing L

XX -amino acid or its analog than the unmodified strain or the wild type

XX strain, or the parental strain of the bacterium. The present sequence

XX encodes the E. coli b2683 protein which has L-amino acid excretion

XX activity.

XX Sequence 336 BP; 62 A; 90 C; 93 G; 91 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.01e-65 Length: 336
 Score: 552.00 Matches: 111
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 100.0% Indels: 0
 DB: 14 Gaps: 0

US-10-073-293A-6 (1-111) x ADZ36196 (1-336)

Qy 1 MetSerTyrGluValLeuLeuGlyLeuLeuValGlyValAlaAenTyrCysPheArg 20
 Db 1 ATGAGCTATGAGGTTCTGCTGCTGGGTTACTAGTTGCGCTGGCGAATATTGCTTCGCG 60
 Qy 21 TyrLeuProLeuArgLeuArgValGlyAsnAlaArgProThrLysArgGlyAlaValGly 40
 Db 61 TATTTGCGCTGCGCTGCTGCTGGGTAAATGCGCCCAACCAACGTGGCGCGTAGGT 120
 Qy 41 IleLeuLeuAspThrIleGlyIleAlaSerIleCysAlaLeuLeuValSerThrAla 60
 Db 121 ATTTTGTCTGCACACCATTTGCTGCTGCATATGCGCTCTGCTGGTTGTTCTTACCGCA 180
 Qy 61 ProGluValMethHisaspThrArgPheValProThrLeuValGlyPheAlaValLeu 80
 Db 181 CCAGAAAGTGATGCACGATACACGCGTTTCGTGCCACCGCTGCTGCGGTACTG 240
 Qy 81 GlyAlaSerPheTyrLysThrArgSerIleIleIleProThrLeuLeuSerAlaLeuAla 100
 Db 241 GGTGCCAGTTTCTATAAACACGACGATTCATCCCAACACTGCTTAGTGGCGCTGCC 300
 Qy 101 TyrGlyLeuAlaTrpLysValMetAlaIleIle 111
 Db 301 TATGGCTCGCTGGAAGTGATGCCGATTATA 333

RESULT 4

AAK70628

ID AAK70628 standard; DNA; 314 BP.

XX AC AAK70628;

XX 06-NOV-2001 (first entry)

XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:25440.

XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

XX cytostatic; gene therapy; vaccine; metastasis; ds.

XX Homo sapiens.

XX WO200157182-A2.

XX 09-AUG-2001.

XX 17-JAN-2001; 2001WO-US001354.

XX 31-JAN-2000; 2000US-0179065P.

XX 04-FEB-2000; 2000US-0180628P.

XX 24-FEB-2000; 2000US-0184664P.

XX 02-MAR-2000; 2000US-0186350P.

XX 16-MAR-2000; 2000US-0189874P.

XX 17-MAR-2000; 2000US-0190076P.

XX 18-APR-2000; 2000US-0198123P.

XX 19-MAY-2000; 2000US-0205515P.

XX 07-JUN-2000; 2000US-0209467P.

XX 28-JUN-2000; 2000US-0214886P.

XX 30-JUN-2000; 2000US-0215135P.

XX 07-JUL-2000; 2000US-0216647P.

XX 07-JUL-2000; 2000US-0216880P.

XX 11-JUL-2000; 2000US-0217487P.

XX 11-JUL-2000; 2000US-0217496P.

XX 14-JUL-2000; 2000US-0218290P.

XX 26-JUL-2000; 2000US-0220963P.

XX 26-JUL-2000; 2000US-0220964P.

PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225226P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234221P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.

PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides, useful for preventing, diagnosing and/or treating cancers and metastasis.

Disclosure; SEQ ID NO 25440; 3071pp + Sequence Listing; English.

AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I) amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic activity, and can be used in gene therapy and vaccine production. (I) proteins and polynucleotides may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate (I) expression. For example, they may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of (I) by expressing inactive proteins or to supplement the patients own production of (I). Additionally, (I) polynucleotides may be used to produce the secreted (I), by inserting the nucleic acids into a host cell and culturing the cell to express the protein. (I) proteins and polynucleotides may be used to prevent, diagnose and treat immune/haematopoietic-related diseases, especially cancers and cancer metastases of haematopoietic-derived cells. AAK64703 to AAK87694 represent human immune/haematopoietic antigen genomic sequences from the present invention. AAK54942 to AAK54950 and AAM82169

CC represent sequences used in the exemplification of the present invention
XX Sequence 314 BP; 69 A; 86 C; 74 G; 85 T; 0 U; 0 Other;
SQ

Alignment Scores:
Pred. No.: 3,74e-48 Length: 314
Score: 423.00 Matches: 86
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 76.6% Indels: 0
DB: 4 Gaps: 0

US-10-073-293A-6 (1-111) x AAK70628 (1-314)

Qy 26 LeuArgValGlyAsnAlaAArgProThrLysArgGlyAlaValGlyIleLeuLeuAspThr 45
Db 3 CTGGGTGGGTAAATGCGCGCCCAACCAAGTGGCGGTAGTATTTGCTGCACAC 62
Qy 46 IleGlyIleAlaSerIleCysAlaLeuLeuValValSerThrAlaProGluValMetHis 65
Db 63 ATTGGCATCGCTCGATATGCGCTCTGCTGGTTGCTCTACCGCACCAAGTGTATGCAC 122
Qy 66 AspThrArgArgPheValProThrLeuValGlyPheAlaValLeuGlyAlaSerPheTyr 85
Db 123 GATACACGCGCTTTCGTGGCCACACGCTGTCGGCTTCGCGTACTGCGGTCAGTTCTAT 182
Qy 86 LysThrArgSerIleIleProThrLeuLeuSerAlaLeuAlaTyrGlyLeuAlaTyr 105
Db 183 AAAACAGCAGCATATATCAACCAACTGCTTAGTGGCTGGCTATGGCTGCGCTGG 242
Qy 106 LysValMetAlaIleIle 111
Db 243 AAAGTGATGCGATTATA 260

RESULT 5

ACH97252
ID ACH97252 standard; DNA; 372 BP.

XX AC ACH97252;

XX DT 29-JUL-2004 (first entry)

XX DE Klebsiella pneumoniae polynucleotide seqid 3047.

XX KW Recombinant expression vector; transcription regulatory element;

XX KW Klebsiella pneumoniae protein; antibacterial; Vaccine; gene; ds.

XX OS Klebsiella pneumoniae.

XX PN US6610836-B1.

XX PD 26-AUG-2003.

XX PF 27-JAN-2000; 2000US-00489039.

XX PR 29-JAN-1999; 99US-0117747P.

XX PA (GENO-) GENOME THERAPEUTICS CORP.

XX PI Breton GL, Osborne M;

XX DR WPI; 2003-895346/82.

XX DR P-PSDB; ABO63701.

XX PT New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for
PT preparing a vaccine composition against Klebsiella pneumoniae.

XX PS Disclosure; SEQ ID NO 3047; 932pp; English.

XX CC The invention describes a new isolated nucleic acid encoding a Klebsiella
CC pneumoniae polypeptide. Also described are: a recombinant expression
CC vector comprising the nucleic acid, operably linked to a transcription
CC regulatory element; and a cell comprising the recombinant expression

CC vector. The nucleic acid is useful for preparing a vaccine composition
CC against Klebsiella pneumoniae. This sequence encodes a Klebsiella
CC pneumoniae polypeptide of the invention

SQ Sequence 372 BP; 50 A; 123 C; 113 G; 86 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2,55e-45 Length: 372
Score: 403.00 Matches: 77
Percent Similarity: 84.1% Conservative: 13
Best Local Similarity: 72.0% Mismatches: 17
Query Match: 73.0% Indels: 0
DB: 11 Gaps: 0

US-10-073-293A-6 (1-111) x ACH97252 (1-372)

Qy 1 MetSerTyrGluValLeuLeuLeuGlyLeuLeuValGlyValAlaAsnTyrCysPheArg 20
Db 37 ATGAACAGTGAAGTCCCTTCTGCTCGGGGTGATAGTCGGCGCGTCAACTATCTCTTTCG 96
Qy 21 TyrLeuProLeuArgLeuArgValGlyAsnAlaArgProThrLysArgGlyAlaValGly 40
Db 97 TATCTTCTCTGCTGTTTGGCGCGCCAGCTGCTCTGCCCGCCCGCGCTGCTGCTGCTG 156
Qy 41 IleLeuLeuAspThrIleGlyIleAlaSerIleCysAlaLeuLeuValSerThrAla 60
Db 157 GTGCTGCTGATACTATTGGCATCGCTCGATTTGGCGCTGCTGCTGCTGCTGCTGCTG 216
Qy 61 ProGluValMetHisAspThrArgArgPheValProThrLeuValGlyPheAlaValLeu 80
Db 217 CCGGAGATCCTCGCGCATGCCATCGCTGCGCGCCAGCTGACCGCTTCTCTGCTGCTG 276
Qy 81 GlyAlaSerPheTyrLysThrArgSerIleIleIleProThrLeuLeuSerAlaLeuAla 100
Db 277 GGGGCGCCCTCTGGAAACCCGAGCATATTGTCGCCACCATTTGCTGCTGCTGCTGCTG 336
Qy 101 TyrGlyLeuAlaTyrLysVal 107
Db 337 TATGGTCTGGCATGGAAATC 357

RESULT 6

AAK57571

ID AAK57571 standard; cDNA; 238 BP.

XX AC AAK57571;

XX DT 06-NOV-2001 (first entry)

XX DE Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:2631.

XX KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ss.

XX OS Homo sapiens.

XX PN WO200157182-A2.

XX PD 09-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US001354.

XX PR 31-JAN-2000; 2000US-0179065P.

XX PR 04-FEB-2000; 2000US-0180628P.

XX PR 24-FEB-2000; 2000US-0184564P.

XX PR 02-MAR-2000; 2000US-0186350P.

XX PR 16-MAR-2000; 2000US-0189874P.

XX PR 17-MAR-2000; 2000US-0190076P.

XX PR 18-APR-2000; 2000US-0198123P.

XX PR 19-MAY-2000; 2000US-0205515P.

XX PR 07-JUN-2000; 2000US-0209467P.

XX PR 28-JUN-2000; 2000US-0214886P.

XX PR 30-JUN-2000; 2000US-0215135P.

XX PR 07-JUL-2000; 2000US-0216647P.

PR	07-JUL-2000;	2000US-0216880P.	PR	20-OCT-2000;	2000US-0241809P.
PR	11-JUL-2000;	2000US-0217487P.	PR	20-OCT-2000;	2000US-0241826P.
PR	11-JUL-2000;	2000US-0217496P.	PR	01-NOV-2000;	2000US-0244617P.
PR	14-JUL-2000;	2000US-0218290P.	PR	08-NOV-2000;	2000US-0246474P.
PR	26-JUL-2000;	2000US-0220963P.	PR	08-NOV-2000;	2000US-0246475P.
PR	26-JUL-2000;	2000US-0220964P.	PR	08-NOV-2000;	2000US-0246476P.
PR	14-AUG-2000;	2000US-0224518P.	PR	08-NOV-2000;	2000US-0246477P.
PR	14-AUG-2000;	2000US-0224519P.	PR	08-NOV-2000;	2000US-0246478P.
PR	14-AUG-2000;	2000US-0225213P.	PR	08-NOV-2000;	2000US-0246523P.
PR	14-AUG-2000;	2000US-0225214P.	PR	08-NOV-2000;	2000US-0246524P.
PR	14-AUG-2000;	2000US-0225266P.	PR	08-NOV-2000;	2000US-0246525P.
PR	14-AUG-2000;	2000US-0225267P.	PR	08-NOV-2000;	2000US-0246526P.
PR	14-AUG-2000;	2000US-0225268P.	PR	08-NOV-2000;	2000US-0246527P.
PR	14-AUG-2000;	2000US-0225270P.	PR	08-NOV-2000;	2000US-0246528P.
PR	14-AUG-2000;	2000US-0225447P.	PR	08-NOV-2000;	2000US-0246532P.
PR	14-AUG-2000;	2000US-0225757P.	PR	08-NOV-2000;	2000US-0246609P.
PR	14-AUG-2000;	2000US-0225758P.	PR	08-NOV-2000;	2000US-0246610P.
PR	14-AUG-2000;	2000US-0225759P.	PR	08-NOV-2000;	2000US-0246611P.
PR	18-AUG-2000;	2000US-0226279P.	PR	08-NOV-2000;	2000US-0246613P.
PR	22-AUG-2000;	2000US-0226661P.	PR	17-NOV-2000;	2000US-0249207P.
PR	22-AUG-2000;	2000US-0226868P.	PR	17-NOV-2000;	2000US-0249208P.
PR	22-AUG-2000;	2000US-0227182P.	PR	17-NOV-2000;	2000US-0249209P.
PR	23-AUG-2000;	2000US-0227182P.	PR	17-NOV-2000;	2000US-0249210P.
PR	30-AUG-2000;	2000US-0228924P.	PR	17-NOV-2000;	2000US-0249211P.
PR	01-SEP-2000;	2000US-0229287P.	PR	17-NOV-2000;	2000US-0249212P.
PR	01-SEP-2000;	2000US-0229343P.	PR	17-NOV-2000;	2000US-0249213P.
PR	01-SEP-2000;	2000US-0229344P.	PR	17-NOV-2000;	2000US-0249214P.
PR	01-SEP-2000;	2000US-0229345P.	PR	17-NOV-2000;	2000US-0249215P.
PR	05-SEP-2000;	2000US-0229509P.	PR	17-NOV-2000;	2000US-0249216P.
PR	05-SEP-2000;	2000US-0229513P.	PR	17-NOV-2000;	2000US-0249217P.
PR	06-SEP-2000;	2000US-0230437P.	PR	17-NOV-2000;	2000US-0249218P.
PR	06-SEP-2000;	2000US-0230438P.	PR	17-NOV-2000;	2000US-0249244P.
PR	08-SEP-2000;	2000US-0231242P.	PR	17-NOV-2000;	2000US-0249245P.
PR	08-SEP-2000;	2000US-0231243P.	PR	17-NOV-2000;	2000US-0249265P.
PR	08-SEP-2000;	2000US-0231244P.	PR	17-NOV-2000;	2000US-0249265P.
PR	08-SEP-2000;	2000US-0231413P.	PR	17-NOV-2000;	2000US-0249297P.
PR	08-SEP-2000;	2000US-0231414P.	PR	17-NOV-2000;	2000US-0249299P.
PR	08-SEP-2000;	2000US-0232080P.	PR	17-NOV-2000;	2000US-0249300P.
PR	08-SEP-2000;	2000US-0232081P.	PR	01-DEC-2000;	2000US-0250160P.
PR	12-SEP-2000;	2000US-0232196P.	PR	01-DEC-2000;	2000US-0250391P.
PR	14-SEP-2000;	2000US-0232397P.	PR	05-DEC-2000;	2000US-0251030P.
PR	14-SEP-2000;	2000US-0232398P.	PR	05-DEC-2000;	2000US-0251988P.
PR	14-SEP-2000;	2000US-0232399P.	PR	05-DEC-2000;	2000US-0256719P.
PR	14-SEP-2000;	2000US-0232400P.	PR	06-DEC-2000;	2000US-0251479P.
PR	14-SEP-2000;	2000US-0232401P.	PR	08-DEC-2000;	2000US-0251856P.
PR	14-SEP-2000;	2000US-0233063P.	PR	08-DEC-2000;	2000US-0251868P.
PR	14-SEP-2000;	2000US-0233064P.	PR	08-DEC-2000;	2000US-0251869P.
PR	14-SEP-2000;	2000US-0233065P.	PR	08-DEC-2000;	2000US-0251989P.
PR	21-SEP-2000;	2000US-0234223P.	PR	08-DEC-2000;	2000US-0251990P.
PR	21-SEP-2000;	2000US-0234274P.	PR	11-DEC-2000;	2000US-0254097P.
PR	25-SEP-2000;	2000US-0234997P.	PR	05-JAN-2001;	2001US-0259678P.
PR	25-SEP-2000;	2000US-0234998P.	PR		
PR	26-SEP-2000;	2000US-0235484P.	PR		
PR	27-SEP-2000;	2000US-0235834P.	PR		
PR	27-SEP-2000;	2000US-0235835P.	PR		
PR	29-SEP-2000;	2000US-0236327P.	PR		
PR	29-SEP-2000;	2000US-0236367P.	PR		
PR	29-SEP-2000;	2000US-0236368P.	PR		
PR	29-SEP-2000;	2000US-0236369P.	PR		
PR	29-SEP-2000;	2000US-0236370P.	PR		
PR	02-OCT-2000;	2000US-0236802P.	PR		
PR	02-OCT-2000;	2000US-0237037P.	PR		
PR	02-OCT-2000;	2000US-0237038P.	PR		
PR	02-OCT-2000;	2000US-0237039P.	PR		
PR	02-OCT-2000;	2000US-0237040P.	PR		
PR	13-OCT-2000;	2000US-0239935P.	PR		
PR	13-OCT-2000;	2000US-0239937P.	PR		
PR	20-OCT-2000;	2000US-0240960P.	PR		
PR	20-OCT-2000;	2000US-0241221P.	PR		
PR	20-OCT-2000;	2000US-0241785P.	PR		
PR	20-OCT-2000;	2000US-0241786P.	PR		
PR	20-OCT-2000;	2000US-0241787P.	PR		
PR	20-OCT-2000;	2000US-0241808P.	PR		

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-483426/52.

P-PSDB; AAM84790.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides, useful for preventing, diagnosing and/or treating cancers and metastasis.

Claim 1; SEQ ID NO 2631; 3071pp + Sequence Listing; English.

AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I) amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic activity, and can be used in gene therapy and vaccine production. (I) proteins and polynucleotides may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate (I) expression. For example, they may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of (I) by expressing inactive proteins or to supplement the patients own production of (I). Additionally, (I)

CC polynucleotides may be used to produce the secreted (I), by inserting the
 CC nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent the
 CC diagnosis and treat immune/haematopoietic-related diseases, especially
 CC to AA87694 represent human immune/haematopoietic antigen genomic
 CC sequences from the present invention. AA87694 to AA87695 and AA87696
 CC represent sequences used in the exemplification of the present invention
 XX
 SQ Sequence 238 BP; 43 A; 71 C; 61 G; 60 T; 0 U; 3 Other;

Alignment Scores:
 Pred. No.: 6.03e-35 Length: 238
 Score: 325.00 Matches: 73
 Percent Similarity: 93.7% Conservative: 1
 Best Local Similarity: 92.4% Mismatches: 4
 Query Match: 58.9% Indels: 2
 DB: 4 Gaps: 0

US-10-073-293A-6 (1-111) x AA87571 (1-238)

QY 26 LeuArgValGlyAenAlaArgProThrLysArgGlyAlaValGlyIleLeuLeuAspThr 45
 DB 3 CTGGCTGGGTAAATGCGCGCCCAACCAACGAGCGCGGTAGGTATTTCTCGACACC 62
 QY 46 IleGlyIleAlaSerIleCysAlaLeuLeuValValSerThrAlaProGluValMetHis 65
 DB 63 ATTGGCATCG-TGATATGCGCTCTGCTGTTGCTCTACCGCACCAAGATGTCAC 121
 QY 66 AspThrArgGlyPheAlaValGlyIleLeuValGlyPheAlaValSerPheTyr 85
 DB 122 GATACACGCGTTTCGCGCCACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTAT 181
 QY 86 LysThrArgSerIleIleIleProThrLeuLeuSerAla-LeuAlaTyrGlyLeu 103
 DB 182 AAACACGACGATATCATCCACACTGGTAGTGGCTTGGCCATGAGGCTT 236

RESULT 7
 ACF69126
 ID ACF69126 standard; DNA; 339 BP.

XX ACF69126;

XX 20-NOV-2003 (first entry)

XX Photorhabdus luminescens nucleotide sequence #7593.

XX Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
 KW detection; food; gene expression; plant; animal; microorganism; toxin;
 KW antibiotic; biopesticide; virulence factor; disease model; plague;
 KW whooping cough; gene; ds.

XX Photorhabdus luminescens.

XX WO200294867-A2.

XX 28-NOV-2002.

XX 07-FEB-2002; 2002WO-IB003040.

XX 07-FEB-2001; 2001FR-00001659.

XX (INSP) INST PASTEUR.

XX (CNRS) CNRS CENT NAT RECH SCI.

XX Duchaud E, Taourit S, Glaser P, Frangeul L, Kunat F, Danchin A;
 PI Buchrieser C;

XX WPI; 2003-148459/14.

XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
 PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
 XX

PS Claim 2; SEQ ID NO 7593; 1205pp; French.
 XX The invention relates to the isolation of genes and their encoded
 CC proteins from Photorhabdus luminescens. The isolated sequences are
 CC sources of probes and primers for detecting the genome of P. luminescens
 CC and related species; to study polymorphisms; for gene analysis and for
 CC detection/amplification of the genes. Antibodies (Ab) raised against the
 CC polypeptides encoded by the genes are used for detection/identification
 CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
 CC carry a gene-containing vector are used to select compounds that
 CC modulate, regulate, induce or inhibit expression of the genes in plants,
 CC animals or microorganisms other than P. luminescens and are able to alter
 CC response or sensitivity to toxins and antibiotics produced by P.
 CC luminescens. Cells transformed to express the genes are useful for
 CC recombinant production of the proteins, particularly toxins and
 CC antibacterials useful as insecticides, bactericides and fungicides. The
 CC genes, proteins, vectors containing the genes and Ab are also useful
 CC therapeutically (to treat microbial infection by bacteria or fungi that
 CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
 CC biopesticides. Other uses of the genes and the proteins are as virulence
 CC factors and for identifying targets of human diseases for which P.
 CC luminescens is a model (particularly plague and whooping cough). This
 CC sequence represents one of the isolated P. luminescens genes
 XX

SQ Sequence 339 BP; 99 A; 66 C; 62 G; 112 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 2e-27 Length: 339
 Score: 271.50 Matches: 54
 Percent Similarity: 72.0% Conservative: 23
 Best Local Similarity: 50.5% Mismatches: 25
 Query Match: 49.2% Indels: 5
 DB: 10 Gaps: 2

US-10-073-293A-6 (1-111) x ACF69126 (1-339)

QY 4 GluValLeuLeuLeuGlyLeuValGlyValAlaAsnTyrCysPheArgTyrLeuPro 23
 DB 13 AAGATTTTGGTATGGAGCTATTGTTGGGTAGCTAACTTTTCATTTCTGCTATCGCA 72
 QY 24 LeuArgLeuArgValGlyAsnAlaArgProThr-----LysArgGlyAlaValGly 40
 DB 73 CTACGATTT-----GGGAAAGCAGCAATCTGCCGCGAGAAAGCTGGAAACAAAGC 126
 QY 41 IleLeuLeuAspThrIleGlyIleAlaSerIleCysAlaLeuLeuValSerThrAla 60
 DB 127 ATTATCTTGACAGTATGGTATGTCATTCATTTGTTTCTTCTATCATCGTATCAGGTGTA 186
 QY 61 ProGluValMetHisAspThrArgPheValProThrLeuValGlyPheAlaValLeu 80
 DB 187 CCTGATGTGATGAGAGAAAGTCAAAACTACTTCTTACCTCATAGGTTGCTGACCATC 246
 QY 81 GlyAlaSerPheTyrLysThrArgSerIleIleIleProThrLeuLeuSerAlaLeuAla 100
 DB 247 TGTTTAGTCTTTTACAAAACAAAGCAATATTATATCGCAACTATTGTCGCGCACTGCT 306
 QY 101 TyrGlyLeuAlaTrpLysVal 107
 DB 307 TTTGGACTAATCAATAATA 327

RESULT 8

ACF67367 20

Continuation (21 of 57) of ACF67367 from base 2000001 (Photorhabdus luminescens nucleot

WP Sequence split into 57 fragments LOCUS ACF67367 Accession ACF67367

WP	Fragment Name	Begin	End
WP	ACF67367_00	1	110000
WP	ACF67367_01	100001	210000
WP	ACF67367_02	200001	310000
WP	ACF67367_03	300001	410000
WP	ACF67367_04	400001	510000
WP	ACF67367_05	500001	610000
WP	ACF67367_06	600001	710000
WP	ACF67367_07	700001	810000

WP	ACF67367_08	800001	910000	
WP	ACF67367_09	900001	1010000	
WP	ACF67367_10	1000001	1110000	
WP	ACF67367_11	1100001	1210000	
WP	ACF67367_12	1200001	1310000	
WP	ACF67367_13	1300001	1410000	
WP	ACF67367_14	1400001	1510000	
WP	ACF67367_15	1500001	1610000	
WP	ACF67367_16	1600001	1710000	
WP	ACF67367_17	1700001	1810000	
WP	ACF67367_18	1800001	1910000	
WP	ACF67367_19	1900001	2010000	
WP	ACF67367_20	2000001	2110000	
WP	ACF67367_21	2100001	2210000	
WP	ACF67367_22	2200001	2310000	
WP	ACF67367_23	2300001	2410000	
WP	ACF67367_24	2400001	2510000	
WP	ACF67367_25	2500001	2610000	
WP	ACF67367_26	2600001	2710000	
WP	ACF67367_27	2700001	2810000	
WP	ACF67367_28	2800001	2910000	
WP	ACF67367_29	2900001	3010000	
WP	ACF67367_30	3000001	3110000	
WP	ACF67367_31	3100001	3210000	
WP	ACF67367_32	3200001	3310000	
WP	ACF67367_33	3300001	3410000	
WP	ACF67367_34	3400001	3510000	
WP	ACF67367_35	3500001	3610000	
WP	ACF67367_36	3600001	3710000	
WP	ACF67367_37	3700001	3810000	
WP	ACF67367_38	3800001	3910000	
WP	ACF67367_39	3900001	4010000	
WP	ACF67367_40	4000001	4110000	
WP	ACF67367_41	4100001	4210000	
WP	ACF67367_42	4200001	4310000	
WP	ACF67367_43	4300001	4410000	
WP	ACF67367_44	4400001	4510000	
WP	ACF67367_45	4500001	4610000	
WP	ACF67367_46	4600001	4710000	
WP	ACF67367_47	4700001	4810000	
WP	ACF67367_48	4800001	4910000	
WP	ACF67367_49	4900001	5010000	
WP	ACF67367_50	5000001	5110000	
WP	ACF67367_51	5100001	5210000	
WP	ACF67367_52	5200001	5310000	
WP	ACF67367_53	5300001	5410000	
WP	ACF67367_54	5400001	5510000	
WP	ACF67367_55	5500001	5610000	
WP	ACF67367_56	5600001	5648894	

Alignment Scores:
 Pred. No.: 7e-24 Length: 110000
 Score: 271.50 Matches: 54
 Percent Similarity: 72.0% Conservative: 23
 Best Local Similarity: 50.5% Mismatches: 25
 Query Match: 49.2% Indels: 5
 DB: 10 Gaps: 2

US-10-073-293A-6 (1-111) x ACF67367_20 (1-110000)

Qy	4	GluValLeuLeuLeuGlyLeuValGlyValAlaAanTyrCyePheArgTyrLeuPro	23
Db	22899	AAGATTTCGTGATGGACTATTGTTGGGTAGCTAACTTTTCATTTCGCTATCTGCCA	22958
Qy	24	LeuArgLeuArgValGlyAsnAlaArgProThr-----LysArgGlyAlaValGly	40
Db	22959	CTACGATTT-----GGGAAGACGCCCAATTCGCCGACAAAGCTGGAAACACAGC	23012
Qy	41	IleLeuLeuAspThrIleGlyIleAlaSerIleCysAlaLeuLeuValValSerThrAla	60
Db	23013	ATTATCCTTGACAGTATTGGTATTGCATCCATTGTTCTTACTCATCGTATCAGGTGTA	23072
Qy	61	ProGluValMetHisAspThrArgPheValProThrLeuValGlyPheAlaValLeu	80

Db	23073	CCTGATGCTGATGAGAGAAAGTCAAAAGTACTTCTTACCCTCATAGTGTGTGACCATC	23132
Qy	81	GlyAlaSerPheTyrLysThrArgSerIleIleIleProThrLeuLeuSerAlaLeuAla	100
Db	23133	TGTTTAGTCTTTTACAAAACAAAGCAAATATATACTCGCAACACTATTTGGCGCACTGCTT	23192
Qy	101	TyGlyLeuAlaTrpLysVal	107
Db	23193	TTTGAGCTAACATTCAAAATA	23213
RESULT 9			
ACF65382			
ID	ACF65382 standard; DNA; 243072 BP.		
XX	ACF65382;		
XX	20-NOV-2003 (first entry)		
DE	Photorhabdus luminescens nucleotide sequence #35.		
XX	Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;		
KW	detection; food; gene expression; plant; animal; microorganism; toxin;		
KW	antibiotic; biopesticide; virulence factor; disease model; plague;		
XX	whooping cough; gene; ds.		
OS	Photorhabdus luminescens.		
XX	WO200294867-A2.		
XX	28-NOV-2002.		
PF	07-FEB-2002; 2002WO-IB003040.		
XX	07-FEB-2001; 2001FR-00001659.		
XX	(INSP) INST PASTEUR.		
PA	(CNRS) CNRS CENT NAT RECH SCI.		
XX	Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;		
PI	Buchrieser C;		
XX	WPI; 2003-148459/14.		
XX	Genomic sequence of Photorhabdus luminescens and encoded polypeptides,		
PT	useful e.g. as therapeutic antimicrobials and agricultural pesticides.		
XX	Claim 1; SEQ ID NO 35; 1205pp; French.		
PS			
XX			
CC	The invention relates to the isolation of genes and their encoded		
CC	proteins from Photorhabdus luminescens. The isolated sequences are		
CC	sources of probes and primers for detecting the genome of P. luminescens		
CC	and related species; to study polymorphisms; for gene analysis and for		
CC	detection/amplification of the genes. Antibodies (Ab) raised against the		
CC	polypeptides encoded by the genes are used for detection/identification		
CC	of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that		
CC	carry a gene-containing vector are used to select compounds that		
CC	modulate, regulate, induce or inhibit expression of the genes in plants,		
CC	animals or microorganisms other than P. luminescens and are able to alter		
CC	response or sensitivity to toxins and antibiotics produced by P.		
CC	luminescens. Cells transformed to express the genes are useful for		
CC	recombinant production of the proteins, particularly toxins and		
CC	antibacterials useful as insecticides, bactericides and fungicides. The		
CC	genes, proteins, vectors containing the genes and Ab are also useful		
CC	therapeutically (to treat microbial infection by bacteria or fungi that		
CC	are sensitive to P. luminescens-encoded toxins or antibiotics) and as		
CC	biopesticides. Other uses of the genes and the proteins are as virulence		
CC	factors and for identifying targets of human diseases for which P.		
CC	luminescens is a model (particularly plague and whooping cough). This		
XX	sequence represents one of the isolated P. luminescens genes		
XX	Sequence 243072 BP; 70273 A; 57885 C; 49427 G; 65487 T; 0 U; 0 Other;		

PT Escherichia coli K12, useful for detecting gene expression and expression
PT patterns.
XX Claim 3; Page 1834; 2004pp; German.
XX This invention describes a novel biochip comprising probe spots, each
CC containing many identical probes. The probes are nucleotide sequences of
CC 30-80 bases, are prepared ex situ from synthetic oligonucleotides and at
CC least one includes a segment of at least 20 bases identical with, or
CC complementary to, a segment of an open reading frame (orf) of Escherichia
CC coli K12. The biochip is used for specific detection of gene expression
CC in K12 and for determining the gene expression pattern, e.g. for
CC diagnostic determination of which E. coli strains are present in the gut,
CC and to determine the effects of e.g. growth media on gene expression. The
CC biochip provides as comprehensive as possible detection of the K12
CC genome, with simultaneous analysis of many different genes with a single
CC device, and comparison of gene expression between K12 and its mutants or
CC other E. coli strains in a single experiment. Apart from qualitative and
CC quantitative information about gene expression, it also allows
CC synthetic oligonucleotides for preparation of probes allows free
CC variation in probe length and ensures high purity (and thus selectivity,
CC reactivity and reproducibility); also synthetic probes are generally
CC shorter than probes prepared by polymerase chain reaction. ACD68731 to
CC ACD81540 represent oligonucleotide probes used with the biochip described
CC in the invention
XX SQ Sequence 100 BP; 15 A; 31 C; 29 G; 25 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 1.69e-13 Length: 100
Score: 164.00 Matches: 33
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 29.7% Indels: 0
DB: 8 Gaps: 0
US-10-073-293A-6 (1-111) x ACD80498 (1-100)
Qy 50 SerTleCysAlaLeuLeuValValSerThrAlaProGluValMetHisAspThrArgArg 69
Db 1 TCGATATGCGCTCTGCTGTTGTCTTACCGCAGAGTGTGTCACGATACACGCGGT 60
Qy 70 PheValProThrLeuValGlyPheAlaValLeuGlyAla 82
Db 61 TTCGTGCCCCACGCTGCTGCGCTTCGCGTACTGGGTGCC 99
RESULT 12
ACD80497
ID ACD80497 standard; DNA; 100 BP.
XX ACD80497;
XX AC
XX 19-SEP-2003 (first entry)
XX DE E. coli K12 MG1655 biochip probe SEQ ID 11773.
XX Biochip; gene expression; gut; diagnostic; detection; probe; ss.
XX Escherichia coli.
XX OS
XX EPI260592-A1.
XX PN
XX 27-NOV-2002.
XX PD
XX 17-MAY-2001; 2001EP-00112179.
XX PF
XX 17-MAY-2001; 2001EP-00112179.
XX PR (MWGB-) MWG-BIOTECH AG.
XX PI Donner H, Drescher B, Huber A, Weber J;
XX (MWGB-) MWG-BIOTECH AG.

DR WPI; 2003-241155/24.
XX Biochip containing probes complementary with open reading frames in
PT Escherichia coli K12, useful for detecting gene expression and expression
PT patterns.
XX Claim 3; Page 1834; 2004pp; German.
XX This invention describes a novel biochip comprising probe spots, each
CC containing many identical probes. The probes are nucleotide sequences of
CC 30-80 bases, are prepared ex situ from synthetic oligonucleotides and at
CC least one includes a segment of at least 20 bases identical with, or
CC complementary to, a segment of an open reading frame (orf) of Escherichia
CC coli K12. The biochip is used for specific detection of gene expression
CC in K12 and for determining the gene expression pattern, e.g. for
CC diagnostic determination of which E. coli strains are present in the gut,
CC and to determine the effects of e.g. growth media on gene expression. The
CC biochip provides as comprehensive as possible detection of the K12
CC genome, with simultaneous analysis of many different genes with a single
CC device, and comparison of gene expression between K12 and its mutants or
CC other E. coli strains in a single experiment. Apart from qualitative and
CC quantitative information about gene expression, it also allows
CC synthetic oligonucleotides for preparation of probes allows free
CC variation in probe length and ensures high purity (and thus selectivity,
CC reactivity and reproducibility); also synthetic probes are generally
CC shorter than probes prepared by polymerase chain reaction. ACD68731 to
CC ACD81540 represent oligonucleotide probes used with the biochip described
CC in the invention
XX SQ Sequence 100 BP; 18 A; 31 C; 24 G; 27 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 5.93e-13 Length: 100
Score: 160.00 Matches: 33
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 29.0% Indels: 0
DB: 8 Gaps: 0
US-10-073-293A-6 (1-111) x ACD80497 (1-100)
Qy 69 ArgPheValProThrLeuValGlyPheAlaValLeuGlyAlaSerPheTyrLysThrArg 88
Db 2 CGTTCGTGCCCCACGCTGCTGCGCTTCGCGTACTGGGTGCCAGTTTCTATAAACACGC 61
Qy 89 SerIleIleIleProThrLeuLeuSerAlaLeuAlaTyr 101
Db 62 AGCATTATCATCCCAACACTGCTTAGTGGCGTGGCCTAT 100
RESULT 13
ACD80499
ID ACD80499 standard; DNA; 100 BP.
XX ACD80499;
XX AC
XX 19-SEP-2003 (first entry)
XX DT E. coli K12 MG1655 biochip probe SEQ ID 11775.
XX DE Biochip; gene expression; gut; diagnostic; detection; probe; ss.
XX KW Escherichia coli.
XX OS
XX EPI260592-A1.
XX PN
XX 27-NOV-2002.
XX PD
XX 17-MAY-2001; 2001EP-00112179.
XX PF
XX 17-MAY-2001; 2001EP-00112179.
XX PR (MWGB-) MWG-BIOTECH AG.
XX PI (MWGB-) MWG-BIOTECH AG.

XX Donner H, Drescher B, Huber A, Weber J;
XX WPI; 2003-241155/24.
XX Biochip containing probes complementary with open reading frames in
XX Escherichia coli K12, useful for detecting gene expression and expression
XX patterns.
XX Claim 3; Page 1834; 2004pp; German.
XX This invention describes a novel biochip comprising probe spots, each
XX containing many identical probes. The probes are nucleotide sequences of
XX 30-80 bases, are prepared ex situ from synthetic oligonucleotides and at
XX least one includes a segment of at least 20 bases identical with, or
XX complementary to, a segment of an open reading frame (orf) of Escherichia
XX coli K12. The biochip is used for specific detection of gene expression
XX in K12 and for determining the gene expression pattern, e.g. for
XX diagnostic determination of which E. coli strains are present in the gut,
XX and to determine the effects of e.g. growth media on gene expression. The
XX biochip provides as comprehensive as possible detection of the K12
XX genome, with simultaneous analysis of many different genes with a single
XX device, and comparison of gene expression between K12 and its mutants or
XX other E. coli strains in a single experiment. Apart from qualitative and
XX quantitative information about gene expression, it also allows
XX measurements of population densities for the various strains. The use of
XX synthetic oligonucleotides for preparation of probes allows free
XX variation in probe length and ensures high purity (and thus selectivity,
XX reactivity and reproducibility); also synthetic probes are generally
XX shorter than probes prepared by polymerase chain reaction. ACD8731 to
XX ACD81540 represent oligonucleotide probes used with the biochip described
XX in the invention

SQ Sequence 100 BP; 18 A; 29 C; 26 G; 27 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2.85e-12 Length: 100
Score: 155.00 Matches: 32
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 28.1% Indels: 0
DB: 8 Gaps: 0

US-10-073-293A-6 (1-111) x ACD80499 (1-100)

Qy 38 AlaValGlyIleLeuLeuAspThrIleGlyIleAlaSerIleCysAlaLeuValVal 57
Db 3 GCGTAGGTATTTTGGTCGACCATTTGGCATCGCTCGATATCGCTCTGCTGTTGTC 62
Qy 58 SerThrAlaProGluValMetHisAspThrArgArg 69
Db 63 TCTACCGCACCAGAGTGATGCACGATACACGCCGT 98

RESULT 14
ACF79478
ID ACF79478 standard; DNA; 172 BP.
XX ACF79478;
XX 18-DEC-2003 (first entry)
XX Escherichia coli multidrug transporter emrR gene promoter.
XX Promoter; emrR gene; multidrug transporter; ds.
XX Escherichia coli.
XX WO2003066853-A1.
XX 14-AUG-2003.
XX 07-FEB-2003; 2003WO-US003902.
XX

PR 08-FEB-2002; 2002US-0355083P.
PR 13-JUN-2002; 2002GB-00013616.
XX (AVET) AVENTIS PHARM INC.
XX Macneil I, Tiong CUY, Brown K;
XX WPI; 2003-731503/69.
XX

Identifying a DNA molecule that encodes, or a protein involved in the
production of, a natural product having bioactivity by incubating the
transformed bacterial cell under conditions that permit the production of
the natural product.

Disclosure; Page 20; Opp; English.

The present sequence is the promoter of the Escherichia coli multidrug
transporter protein gene emrR. The promoter can be used in the method of
the invention for identifying a DNA molecule that encodes a natural
product having bioactivity, or a protein involved in the production of
the natural product. The method involves: (a) providing a transformed a
bacterial cell comprising (i) a gene fusion encoding a reporter protein
associated with a homeostatic promoter which is activated by the natural
product, and (ii) an exogenous DNA that encodes the natural product or
the protein involved in production of the natural product; (b) incubating
the transformed bacterial cell under conditions that permit production of
the natural product; and (c) detecting the presence of the reporter
protein within the transformed bacterial cell

SQ Sequence 172 BP; 49 A; 37 C; 32 G; 54 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 0.00037 Length: 172
Score: 98.00 Matches: 20
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 17.8% Indels: 0
DB: 10 Gaps: 0

US-10-073-293A-6 (1-111) x ACF79478 (1-172)

Qy 92 IleProThrLeuLeuSerAlaLeuAlaTyrGlyLeuAlaTrpLysValMetAlaIleIle 111
Db 2 ATCCCAACACTGCTTAGTGGCTGCTATGGCTCGCTCGAAAGTGATGGCGATTATA 61

RESULT 15
ACA39729
ID ACA39729 standard; DNA; 1593 BP.
XX ACA39729;
XX 19-JUN-2003 (first entry)
XX Prokaryotic essential gene #21386.
XX Antisense; ds; prokaryotic essential gene; cell proliferation;
XX drug design; gene.
XX Mycobacterium leprae.
XX WO200277183-A2.
XX 03-OCT-2002.
XX 21-MAR-2002; 2002WO-US009107.
XX 21-MAR-2001; 2001US-00815242.
XX 06-SEP-2001; 2001US-00948993.
XX 25-OCT-2001; 2001US-0342923P.
XX 08-FEB-2002; 2002US-00072851.
XX 06-MAR-2002; 2002US-0362699P.
XX (ELIT-) ELITRA PHARM INC.
XX

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH,
XX WPI; 2003-029926/02.
DR P-PSDB; ABU35859.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 14; SEQ ID NO 27599; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1593 BP; 239 A; 469 C; 567 G; 318 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 4.61 Length: 1593
Score: 78.00 Matches: 31
Percent Similarity: 40.5% Conservative: 14
Best Local Similarity: 27.9% Mismatches: 32
Query Match: 14.1% Indels: 34
DB: 8 Gaps: 5

US-10-073-293A-6 (1-111) x ACA39729 (1-1593)

QY 4 GluValLeu-----LeuLeuGlyLeuValGlyValAlaAlaAsnTyrCysPheArg 20
Db 166 GAGGTGCTGTCGGGTGATCTGTCGTCATTGGTGGTGCATCCACACGGCGCTCGG 225
QY 21 TyrLeuProLeuArgLeuArgValGlyAsnAlaArgProThrLysArgGlyAlaValGly 40
Db 226 TATCCCGCGTTCGCTC-----GAACGAGGTGCTGTCGCG 261
QY 41 IleLeuLeuAspThrIleGlyIleAlaSerIleCysAlaLeuLeuValValSerThrAla 60
Db 262 GTGCTTACCGAGCTTGAGGGGTCCCGAGCTGACAGCC-----CAGGCCAGCAGCGTG 315
QY 61 ProGluValMetHisAspThrArgArgPheValProThrLeuValGlyPheAlaValLeu 80
Db 316 CCGATACCTCATACATCCGGAACCCGCG-----AGCGTGCTC 351

Qy 81 GlyAlaSerPheTyrLysThrArgSerIleIleIleProThrLeuLeuSerAlaLeuAla 100
Db 352 GGCGGT-----CTTGGCGCGCGGCA 372
Qy 101 TyrGlyLeuAlaTrpLysValMetAlaIleIle 111
Db 373 TATGACATCCCTCCAAACCGGATGACGGTTGTT 405

Search completed: February 16, 2006, 18:26:27
Job time : 350.935 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 16, 2006, 17:12:59 ; Search time 2260.53 Seconds
(without alignments)
2297.407 Million cell updates/sec

Title: US-10-073-293A-6

Perfect score: 552

Sequence: 1 MSVEVLGLLVGVNYCFR.....IPTLLSALAYGLAWKVMAIL 111

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/abs/ABSSWEB.spool/US10073293/runat.15022006.120736.1008/app.query.fasta_1
-DB=EST -QPMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abs04
-USER=US10073293 @CGN 1.1 6731 @runat.15022006.120736.1008 -NCPU=6 -ICPU=3
-NO MMAP -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_hic.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_est7.*
9: gb_gss1.*
10: gb_gss2.*
11: gb_gss3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	91.5	16.6	754	2 BG721873	BG721873 602694488
2	86.5	15.7	511	2 BG881410	BG881410 sae1g11.
3	86.5	15.7	776	8 DR425028	DR425028 naw13g11.
4	86.5	15.7	897	5 BU375062	BU375062 603588638
5	86.5	15.7	916	5 BU377599	BU377599 603812270
6	86.5	15.7	958	5 BU274235	BU274235 603532828
7	85.5	15.5	651	7 CN847401	CN847401 PG07029C0

8	85	15.4	755	7	CV111390
9	83.5	15.1	547	8	DN993155
c 10	83	15.0	584	3	BI972603
11	82.5	14.9	431	2	BB844978
c 12	82.5	14.9	717	1	AJ448384
c 13	82.5	14.9	728	8	DR426972
c 14	82.5	14.9	730	1	AJ446899
c 15	82.5	14.9	788	1	AJ446332
c 16	82.5	14.9	811	1	AJ452083
c 17	82.5	14.9	827	1	AJ455413
c 18	82.5	14.9	958	5	BU421186
19	82	14.9	534	1	AL681632
20	82	14.9	655	1	AL849287
21	82	14.9	661	1	AL856042
22	82	14.9	681	9	BZ088402
23	82	14.9	706	7	CN102523
24	82	14.9	813	1	AL965519
25	82	14.9	836	7	CN095349
26	82	14.9	843	8	CX442865
27	82	14.9	847	9	BH465150
28	82	14.9	1701	10	AY408608
c 29	82	14.9	2005	2	BF034156
30	82	14.9	2160	4	AK016870
31	81.5	14.8	432	6	CB385955
32	81	14.7	767	7	CK599023
c 33	80.5	14.6	594	3	BJ675006
34	80.5	14.6	651	2	BF337569
35	80.5	14.6	818	2	BG720995
36	80.5	14.6	947	5	BQ652127
c 37	79.5	14.4	499	11	TA369G01P
c 38	79.5	14.4	667	6	CD475939
c 39	79.5	14.4	770	2	BF121899
40	79	14.3	864	7	CO733249
41	78.5	14.2	590	8	DN956994
42	78.5	14.2	591	7	CN303248
43	78.5	14.2	738	7	CN303263
c 44	78.5	14.2	918	10	CG004411
45	78.5	14.2	944	7	CO583414

ALIGNMENTS

RESULT 1

BG721873

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BG721873 754 bp mRNA linear EST 08-MAY-2001
602694488F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4826757 5',
mRNA sequence.

BG721873 GI:14001060

EST.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

1 (bases 1 to 754)

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs@mail.nih.gov

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.

cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki

Toshiyuki and Piero Carninci (RIKEN)

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

<http://image.lnl.gov>

Plate: LLM10741 row: p column: 22

High quality sequence stop: 691.

Location/Qualifiers

1..754

FEATURES

source

Qy 80 LeuGlyAla-----SerPheTyrLysThrArgSerIleIleIleProThrLeuLeuSer 97
 Db 404 CTTCAAGCTCTCACTCGTACTCCAGACTCAGAGTATGATCTTTCCATGTTTTCAGC 463
 Qy 98 AlaLeu 99
 Db 464 TCAATC 469

RESULT 3
 DR425028/c
 LOCUS
 DEFINITION
 nawi3911.y1 Chicken eye (hatched). Unnormalized (new) Gallus gallus
 cDNA clone nawi3911 5', mRNA sequence.
 DR425028
 DR425028.1 GI:68327044
 EST.
 SOURCE
 Gallus gallus (chicken)
 ORGANISM
 Gallus gallus
 Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 1 (bases 1 to 776)
 Wistow, G., Peterson, K. and McMurtry, J.
 NBI Bank analysis of 15day post-hatched chicken eye
 Unpublished (2005)
 Contact: Wistow G
 Section on Molecular Structure and Function
 National Eye Institute
 6/331, NIH, Bethesda, MD 20892-2740, USA
 Tel: 301 402 3452
 Fax: 301 496 0078
 Email: graeme@helix.nih.gov
 Plate: 13 row: 9 column: 11
 Seq primer: Universal M13 Reverse.

FEATURES
 source
 1..776
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /db_xref="taxon:9031"
 /clone="nawi3911"
 /tissue_type="Whole eye"
 /dev_stage="15d post-hatched"
 /lab_host="EMDH10B"
 /clone_lib="Chicken eye (hatched). Unnormalized (new)"
 /notes="Organ: Eye; Vector: pCMVSPORT6; RNA was extracted from pooled 15day post-hatched chicken eye. A directionally cloned cDNA library in the pCMVSPORT6 vector (Invitrogen) was constructed at Bioserve Biotechnology (Laurel MD) essentially following the protocols of the SuperScript Plasmid System, full details of which are contained in the manufacturer's instruction manual (http://www.lifetech.com/). First strand synthesis was carried out using a Not I primer-adaptor [5'-pGACTAGTCTAGTCGAGCGGCCCT(T)15-3']. cDNA was cloned in Not I/Sal I sites. EST analysis was performed at the NIH Intramural Sequencing Center (NISC). Analyzed data available through http://neibank.nei.nih.gov."

ORIGIN
 Alignment Scores:
 Pred. No.: 2.48 Length: 776
 Score: 86.50 Matches: 38
 Percent Similarity: 45.5% Conservative: 12
 Best Local Similarity: 34.5% Mismatches: 51
 Query Match: 15.7% Indels: 9
 DB: 8 Gaps: 3

US-10-073-293a-6 (1-111) x DR425028 (1-776)

Qy 5 ValLeuLeuGlyLeuValGlyValAlaAsnTyrCysPheArgTyrLeuProLeu 24
 Db 703 GTCTTGCTGTTGCTGCTGTTGTTGTTGTT-----ATCTTGCTGTTGCTGTT 653

Qy 25 ArgLeuArgValGlyValGlyValAlaArgProThrLysArgGlyAlaValGlyIleLeuLeuAsp 44
 Db 652 GCTTTGCTGTTGCTGCTGCTGTTGGCAGCCCACT---GTTGCTTCTTCTGCGCATTTG 596
 Qy 45 ThrIleGlyIleAlaSerIleCysAlaLeuValValSerThrAlaProGluValMet 64
 Db 595 GCCATCGGTGTTGCCATCTTCTCGGCCCATTTTCCCATCATGTTGCCATTTTGTCTCGGT 536
 Qy 65 HisAspThrArgPheValProThrLeuValGlyPheAlaValLeuGlyAlaSerPhe 84
 Db 535 CTGACTGCCCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 476
 Qy 85 TyrLysThrArgSerIleIleIleProThrLeuLeuSerAlaLeuAlaTyr----- 101
 Db 475 GTGGGTGTTTGTGCGCGTATCATCACCCCTGGTGTGCTGCTGCTGCTGCTGCTGCTGCT 416
 Qy 102 -----GlyLeuAlaTrpLysValMetAla 109
 Db 415 GGTGCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 386

RESULT 4
 BU375062/c
 LOCUS
 DEFINITION
 BU375062 897 bp mRNA linear EST 28-NOV-2002
 60358638F1 CSEQHN74 Gallus gallus cDNA clone ChEST54819 5', mRNA sequence.
 ACCESSION
 BU375062
 VERSION
 BU375062.1 GI:25883063
 KEYWORDS
 EST.
 SOURCE
 Gallus gallus (chicken)
 ORGANISM
 Gallus gallus
 Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 1 (bases 1 to 897)
 Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E., Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
 A Comprehensive Collection of Chicken cDNAs
 Curr. Biol. 12 (22), 1965-1969 (2002)
 12445392
 Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology (UMIST)
 PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.

FEATURES
 source
 1..897
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="Compton Line 151"
 /db_xref="taxon:9031"
 /clone="ChEST54819"
 /sex="Female"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_lib="CSEQHN74"
 /note="Organ: kidney + adrenal; Vector: pBluescript II KS(+); Site 1: EcoRI; Site 2: NotI; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunt-ended, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

ORIGIN


```

Db      615  CTTGTTTGGTGGGA  629
|||||
RESULT 8
CV111390
LOCUS
DEFINITION CV111390 755 bp mRNA linear EST 30-AUG-2004
IMAGE:7457639 5', mRNA sequence.
ACCESSION CV111390
VERSION
KEYWORDS
SOURCE
ORGANISM Rattus norvegicus (Norway rat)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 755)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgapbs-remail.nih.gov
Tissue Procurement: Howard Jacobs
cDNA Library Preparation: Express Genomics
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLNL15723 row: d column: 21
High quality sequence stop: 703.
FEATURES
source
1..755
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="IMAGE:7457639"
/tissue_type="testis, pooled"
/lab_host="DH10B Tona"
/clone_lib="NIH MGC 238"
/notes="Organ: testis; Vector: pExpress-1; Site 1: EcorV;
Site 2: NotI; RNA obtained from testis tissue of 8 wk old
animal. Tissues were snap-frozen and kept at -80C before
RNA extraction and purification (Tri-reagent method). cDNA
was primed using oligo-dT primer:
5'-pGACTAGTCTAGATCGAGCGCGCC(T)25-3' and cloned into
the EcorV/NotI sites of pExpress-1. Size-selection >1.4kb
resulted in an average insert size of 1.9 kb. This primary
library is normalized (non-normalized primary library is
NIH MGC 237) and was constructed by Express Genomics
(Frederick, MD)"
ORIGIN
Alignment Scores: 3.74 Length: 755
Pred. No.: 85.00 Matches: 31
Score: 43.9% Conservative: 12
Best Local Similarity: 31.6% Mismatches: 39
Query Match: 15.4% Indels: 16
DB: 7 Gaps: 5
US-10-073-293A-6 (1-111) x CV111390 (1-755)
Qy 18 CysPheArgTyrLeuPro---LeuArgLeuValGlyAsnAlaArgProThrIysArg 36
Db 231 TGTGGAGGCTCAGGCCACGACGACGACGAGTTCCGGCGCTCGGCCCTCGAAGAGA 290
Qy 37 GlyAlaVal-----GlyIleLeuLeuAspThrIleGlyIleAlaSerIle 51
Db 291 GCTGCCATCTTAACATACGAGGCGCTGTATTTCCAGGCTTTCGGAGGACGCGCGC 350

```

```

Qy 52 CysAlaLeuLeuValValSerThrAlaProGluValMetHisAspThrArgAlaGpPheVal 71
Db 351 TGC-----CTCAGCAAGAGTGTGCCCCCATGTAGCCATGGCTACCGCGCTTCTCTA 404
Qy 72 ProThrLeuValGlyPheAlaValLeu-----GlyAlaSerPheTyrLysThrArg 88
Db 405 CCTAGCAAGGACACCGTGTTCATCTTCGACTATGTGGGGCCCTCTTCTTCTCTCAAAAT 464
Qy 89 SerIleIleLeuProThrLeuLeuSerAlaLeuAlaTyrGlyLeuAlaTrpLys 106
Db 465 GCTCTGCTGCTCTCCACCCTTGGCCAG-----ATCCAGTGGAG 503
RESULT 9
LOCUS
DEFINITION DN993155 547 bp mRNA linear EST 17-MAY-2005
Tissue: Human adult whole brain, large insert, pCMV expression
library Homo sapiens cDNA clone TC102327 5' similar to Homo sapiens
cDNA FLJ38307 fis, clone FCBBF3018949, mRNA sequence.
ACCESSION DN993155
VERSION
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 547)
AUTHORS Birkett,C., Cho,J., Gau,Y., Hamer,R., Kelly,S., Kovacs,K., Liu,L.,
Liu,X., Porter,J., Sachs,A., Shu,Y., Sun,Z., Wong,J., Wu,M.,
Zhang,X., Jay,G. and He,W.
TITLE High-throughput cloning of full-length human cDNAs directly from
cDNA libraries optimized for large and rare transcripts
JOURNAL Unpublished (2005)
COMMENT Contact: Kovacs, KF
High Throughput cDNA Cloning
OriGene Technologies, Inc. ( www.oriGene.com )
6 Taft Court, Suite 100, Rockville, MD 20850, USA
Tel: 301 340 3188
Fax: 301 340 8606
Email: cDNA@origene.com
This EST submission is part of an on-going human full-length
cloning project at OriGene Technologies, Inc.
Please contact OriGene for access.
OriGene Technologies, Inc.
6 Taft Ct. Suite 100
Rockville, MD 20850
Tel: (301) 340-3188
http://www.oriGene.com
Seq primer: pCMV6 5prime forward vector primer, OriGene
Technologies Inc.
Location/Qualifiers
1..547
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="TC102327"
/tissue_type="Whole brain"
/clone_lib="Human adult whole brain, large insert, pCMV
expression library"
/notes="Organ: Brain; Vector: pCMV6-XL5; Site 1: EcorI;
Site 2: XhoI/SalI compatible end ligatio; Oligo-dT primed
reverse transcription optimized for large and GC rich mRNA
transcripts, cDNA size selection, optimized ligation for
large inserts into mammalian expression vector, random
clones selected for end sequence verification of
full-length genes"
ORIGIN
Alignment Scores: 3.83 Length: 547
Pred. No.: 83.50 Matches: 34
Score: 40.2% Conservative: 15
Percent Similarity:

```

Best Local Similarity: 27.9% Mismatches: 44
 Query Match: 15.1% Indels: 29
 DB: 8 Gaps: 6

US-10-073-293A-6 (1-111) x DN93155 (1-547)

Qy 7 LeuLeuGlyLeuValGlyValAlaAsnTyrCysPheArgTyrLeuProLeuArgLeu 26
 |||||
 Db 119 CTTTGGGCTTCTCACCTGGGGGGTGGTTCGTGCACACCTCCACCTCTTCCTC 178
 |||||
 Qy 27 ArgValGlyAan-----
 |||||
 Db 179 CGTGTGACGATAGAGCACCATCCAGCAGACCTCCCGGGCGCTTCATCTCCT 238
 |||||
 Qy 34 ThrlysArgGlyAlaValGlyIleLeuLeuAspThrIleGlyIleAlaSer----- 50
 |||||
 Db 239 CGTCTCCAGGCAGCACAC-----AGCCATTGTGGAACTCTCCACCGGTGTACA 286
 |||||
 Qy 51 -----lleCysAlaLeu---LeuValValSerThrAlaProGluValMet----- 64
 |||||
 Db 287 GAACGGTGCCTCTCGTCTCCCTCCACTCAGGACCTCTCAAGTCCCGGATGTGATGGCTCCT 346
 |||||
 Qy 65 ---HisAspThrArgArgPheValProThrLeuValGlyPheAlaValLeuGlyAlaSer 83
 |||||
 Db 347 CAGCATGATCAGGAGAAATCCATGAT-----CTTGTATTCTCTCTTGGGAAGTCC 400
 |||||
 Qy 84 PheTyrIysThrArgSerIleIlePheProThrLeuLeuSerAlaLeuAlaTyrGlyLeu 103
 |||||
 Db 401 TTCTCCATGCTTAACCAAGATCTATATGGCTATAGCACCGACTCTTGGCTCTTGGCTTG 460
 |||||
 Qy 104 AlaTyr 105
 |||||
 Db 461 GCATGG 466

RESULT 10

BI972603/c
 LOCUS
 DEFINITION
 BI972603.1 GI:16347008
 EST.
 Glycine max (soybean)
 Glycine max
 Glycine max (soybean)
 Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.
 1. (bases 1 to 584)
 Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Corvelli, V.,
 Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J.,
 Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M.,
 Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N.,
 Schurk, R., Ritter, S., Kohn, S., Shin, T., Jackson, Y., Cardenas, M.,
 McCann, R., Waterston, R., and Wilson, R.
 Public Soybean EST Project
 Unpublished (1999)
 Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 When it has been determined, an EST from the other end of this
 clone is listed in the 'Other ESTs on clone' field. Possible
 reversed clone: similarity on wrong strand this clone is available
 through: Biotenetic Services, 801 32nd Ave. Brookings, SD 57006 USA
 (phone: 800 423 4163; email: info@biogeneticservices.com)
 High quality sequence stop: 391.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1. 584

/organism="Glycine max"
 /mol_type="mRNA"
 /cultivar="Williams"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl065-7052"
 /tissue_type="germinating shoots"
 /lab_host="DH10B"
 /clone_lib="Gm-cl065"

/notes="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:
 XhoI; The cDNA library was constructed from mRNA isolated
 germinating shoots of the cultivar Williams. The seeds
 were allowed to germinate for 24 hours prior to being
 cold stressed for 2 days at 4C. Complementary DNA was
 synthesized from mRNA using a primer consisting of a
 poly(dT) sequence with a XhoI restriction site. EcoRI
 adapters were ligated to the blunt-ended cDNA fragments
 directionally cloned into the blunt-ended cDNA fragments
 of the pBluescript vector. The ligated cDNA fragments were
 transformed into DH10B host cells (GibcoBRL). This library
 was constructed in the laboratory of Dr. Randy
 Shoemaker."

ORIGIN

Alignment Scores:
 Pred. No.: 4.84 Length: 584
 Score: 83.00 Matches: 29
 Percent Similarity: 45.7% Conservative: 24
 Best Local Similarity: 25.0% Mismatches: 35
 Query Match: 15.0% Indels: 29
 DB: 3 Gaps: 4

US-10-073-293A-6 (1-111) x BI972603 (1-584)

Qy 23 ProLeuArgLeuArgValGlyAsnAlaArgProThrIysArgGlyAlaValGlyIleLeu 42
 |||||
 Db 532 CCGGTGCTATGGAACCTTTATGTGGCAACCTATGTGGGAAGAGTATAGGAAGTTG 473
 |||||
 Qy 43 LeuAspThrIleGlyIle-----
 |||||
 Db 472 GAAATTACACTTGGTGGCGATTGTGACTGTGACATTGTTGCTCCGATATCTCTGG 413
 |||||
 Qy 49 -----AlaSerIleCysAlaLeuLeuValSerThrAlaProGluValMetHis 65
 |||||
 Db 412 TGTGGATATTCACTGATAAA-ATACTGTCTGCTTAGTCAAGACCTGAGATTCTCAT 354
 |||||
 Qy 66 AspThrArgArgPhe-----ValProThrLeuValGlyPheAlaValLeuGly 81
 |||||
 Db 353 GCAGCTCGTGAGTACTGCATATACCTCATTCAGCTCTATTGGTCATGCTGCTTCAA 294
 |||||
 Qy 82 Ala-----SerPheTyrIysThrArgSerIleIleProThrLeuLeuSerAla--- 98
 |||||
 Db 293 GCCTCACTCGTACTTCCAGACTCAGATGATGATCTTCCCATGGTTTCAGCTCATC 234
 |||||
 Qy 99 -----LeuAlaTyrGlyLeuAlaTyrIysValMetAlaIle 110
 |||||
 Db 233 ACCGCACCTGGTTTGCATGTTCTCTATTGTTGGAGTCTCTGATTGGTT 186
 |||||

RESULT 11

BB844978
 LOCUS
 DEFINITION
 BB844978.1 RIKEN full-length enriched, adult male kidney Mus musculus
 cDNA clone F520002L02 5', mRNA sequence.

ACCESSION BB844978.1 GI:17083353

VERSION BB844978

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridae; Muridae; Muridae; Muridae; Muridae;

REFERENCE 1 (bases 1 to 431)

AUTHORS Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T.,

Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Imotani, K., Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kumada, M., Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Nomasa, R., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, Y., Tomaru, A., Toyota, T., Watanishi, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al. 2001)

[illegible]

US-10-073-293A-6 (1-111) x BB44978 (1-431)

Qy	15	AlaAsnTyrCysPheArgTyrLeuProLeuArgValGlyAsnAlaArgProThr	34
Db	18	AGTGACTTGCCTCGC---TCAGCTTCAGCTTCAGTTCAAGGACGCCCGACC	74
Qy	35	LysArgGlyAlaValGlyIleLeuLeuAspThrIleGlyIleAlaSerIleCysAlaLeu	54
Db	75	ATGAAGACACCGTGAAGGTTCTTCGGAGCTGCTGGTGTGCT-----GCGCTT	125
Qy	55	LeuValValSerThrAlaPro-----GluValMethHisAspThr	67
Db	126	GTCACCATCATCACCCTGCCAATAGTTCTGCTGAGCAAGATGAAGCGGAGCTGCAGC	185
Qy	68	ArgArgPhe-----ValProThr	73
Db	186	CGCAGAAGGTATTCATAGCTGACTATTTAAAGAGTACCTTTCGGGTCAAGTCCCTACT	245
Qy	74	LeuVal---GlyPheAlaValLeuGlyAlaSerPheTyrLysThrArgSerIleIleIle	92
Db	246	CTTTGTGGTGGGTTTCAGACCTTGAATACCTCTACAAACAAG---AGGACAATATCTTGC	302
Qy	93	ProThrLeuLeuSerAlaLeuAlaTyrGlyLeuAlaTrpLysValMetalIle	110
Db	303	TGCTCAATGCTGAACATGGAAACAGCTCCATTTTCTTGGAGACAGTACCTTGG	356

[illegible]

```

ORIGIN
Alignment Scores:
Pred. No.: 7.34
Score: 82.50
Percent Similarity: 44.5%
Best Local Similarity: 33.6%
Query Match: 14.9%
DB: 1
Length: 717
Matches: 37
Conservative: 12
Mismatch: 52
Indels: 9
Gaps: 3
/note="CB mirrored strain"

```

US-10-073-293A-6 (1-111) x AJ448384 (1-717)

Qy 5 ValLeuLeuGlyLeuLeuValcglyvalalaasnTyrCysPheArgTyrLeuProLeu 24
|||||:::|||||

Db 698 GTCATCGCTGGTGAGTCTGTTGGGTGGT-----ATCTGTGGGTGCTGTT 648


```

Qy 21 TyrLeuProLeuArgLeuArgValGlyAsnAlaArgProThrLysArgGlyAlaValGly 40
Db 97 TATCTTCTCGCTTTCGGCGCCGACCGCTCGTCTGCCCGCGCGCGCGT CAGC 156
Qy 41 IleLeuLeuAspThrIleGlyAlaSerIleCysAlaLeuLeuValSerThrAla 60
Db 157 GTGCTGCTGGATATATTGGCATCGCTCGATTTGGCGCTGCTGGTGTCTCCAGCGTG 216
Qy 61 ProGluValMetHisAspThrArgPheValProThrLeuValGlyPheAlaValLeu 80
Db 217 CCGGAGATCTCTCGCGCATGCCATCGCTGGCGCGCAGCTGACCGGCTTCTCGTGTGTG 276
Qy 81 GlyAlaSerPheTyrLysThrArgSerIleIleProThrLeuSerAlaLeuAla 100
Db 277 GGGCGCGCTCTTGGAAACCCGAGCATCATTTGTCGGACCTGCTGTCGGCGTTTGCC 336
Qy 101 TyrGlyLeuAlaTrpLysVal 107
Db 337 TATGCTCTGGCATGGAAATC 357

```

RESULT 2

```

US-09-328-352-2738
; Sequence 2738, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 2738
; LENGTH: 369
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-2738

```

```

Alignment Scores:
Pred. No.: 3.84e-15 Length: 369
Score: 182.50 Matches: 39
Percent Similarity: 63.5% Conservative: 34
Best Local Similarity: 33.9% Mismatches: 35
Query Match: 33.1% Indels: 7
DB: 3 Gaps: 3

```

```

US-10-073-293A-6 (1-111) x US-09-328-352-2738 (1-369)
Qy 1 MetSerTyrGluValLeuLeuLeuGlyLeuLeuValGlyValAlaAsnTyrCysPheArg 20
Db 28 ATGAACCTAGAAATATTATTTGGTCGGCATTAATGTGGGTATGTCTTAATTTTGTCTCAGC 87
Qy 21 TyrLeuPro-----LeuArgLeuArgValGlyAsnAlaArgProThrLysArgGly 37
Db 88 TCGGACCATTTTCGTATACAAAACTACAGGGCAACCAAA-----AGACGTGCG 141
Qy 38 AlaVal-----GlyIleLeuLeuAspThrIleGlyIleAlaSerIleCysAlaLeuLeu 55
Db 142 TCAATCTCGTTTAAATAATCGCTTAGGAAGTATTGGTATATCTGCAATATTAGCGCCATGCTT 201
Qy 56 ValValSerThrAlaProGluValMetHisAspThrArgPheValProThrLeuVal 75
Db 202 GTGGTTCACACATTCACCTTTTACTGAAACCCCAATAAGAGTTTAGCAATGCTTATC 261
Qy 76 GlyPheAlaValLeuGlyAlaSerPheTyrLysThrArgSerIleIleProThrLeu 95
Db 262 GGTATTTTAGTTTAGCAGGCTTTATTTTAAATTTAAGAAATGTTACCAAGCACTTTA 321
Qy 96 LeuSerAlaLeuAlaTrpLysValMetAlaIle 110
Db 322 ACTGCCGAATCGTTTATGGCGCTTATCTATATCTTATTTACCTTTA 366

```

RESULT 3

```

US-09-596-002-10
; Sequence 10, Application US/09596002
; Patent No. 6632636
; GENERAL INFORMATION:
; APPLICANT: Lagace, Robert, E.
; APPLICANT: Patterson, Chandra
; APPLICANT: Berg, Kim, L.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
; FILE REFERENCE: PM-0008-4 US/09/596,002
; CURRENT APPLICATION NUMBER: 2000-06-16
; PRIOR FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: 60/140,121
; PRIOR FILING DATE: 1999-06-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PERL Program
; SEQ ID NO 10
; LENGTH: 19988
; TYPE: DNA
; ORGANISM: M. catarrhalis
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte template ID No. 6632636 10
; PUBLICATION INFORMATION:
US-09-596-002-10

```

Alignment Scores:

```

Pred. No.: 86.1 Length: 19988
Score: 77.50 Matches: 32
Percent Similarity: 40.3% Conservative: 20
Best Local Similarity: 24.8% Mismatches: 36
Query Match: 14.0% Indels: 41
DB: 3 Gaps: 6

```

US-10-073-293A-6 (1-111) x US-09-596-002-10 (1-19988)

```

Qy 9 GlyLeuLeuValGlyValAlaAsnTyrCysPheArgTyr----- 21
Db 2843 GGTGCGCTCATCGGCTTAATCATCAGTGTTCGACCGCTTTTACCATCTTTGCCAACATTGAT 2902
Qy 22 -----LeuProLeuArgLeuArgValGlyAsnAlaArgProThrLys 35
Db 2903 TTGATTTTAGATGCTGACTTACCCTCATCAATTAATGTGGCAATGTTTCATCCATGG--- 2959
Qy 36 ArgGlyAlaValGlyIleLeuLeuAspThrIleGlyIleAlaSerIleCysAlaLeuLeu 55
Db 2960 -----CTTGGCACTTTGATGAGTTTATCATCTTTTGCC-----ATG 2995
Qy 56 ValValSerThrAlaProGluValMetHis-----Asp 66
Db 2996 ATTTTAAATACCGCCATCAGTTTATATCTCGCTGGCCAGCGCTTTTTCAGGTGATGAC 3055
Qy 67 ThrArgArgPheValProThrLeuValGlyPheAlaValLeuGly----- 81
Db 3056 AACGAGCGCTTAAATAGATATATGATGGCTTGTTTGGTCGCTTTATCTTAAGCTTT 3115
Qy 82 AlaSerPheTyrLysThrArgSerIleIleIleProThr-----LeuLeu 96
Db 3116 GCAGGTTTAAATAAACGTGCTCAATCATGTTCCCATCATCTGGCTACATTTGGTATGTTG 3175
Qy 97 SerAlaLeuAlaTyrGlyLeuAlaTrp 105
Db 3176 CTTCTTCTCGTGTGTTGTTGTCATGG 3202

```

RESULT 4

```

US-09-711-164-57
; Sequence 57, Application US/09711164
; Patent No. 6589738
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERET

```

```
; FILE REFERENCE: ELITRA.008A
; CURRENT APPLICATION NUMBER: US/09/711.164
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/164415
; PRIOR FILING DATE: 1999-11-9
; NUMBER OF SEQ ID NOS: 469
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 57
; LENGTH: 330
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-711-164-57
```

```
Alignment Scores:
Pred. No.: 0.313 Length: 330
Score: 77.00 Matches: 24
Percent Similarity: 41.9% Conservative: 20
Best Local Similarity: 22.9% Mismatches: 37
Query Match: 13.9% Indels: 24
DB: 3 Gaps: 3
```

US-10-073-293A-6 (1-111) x US-09-711-164-57 (1-330)

```
Qy 24 LeuArgLeuValGlyAsnAlaArgProThrLysArgGlyAlaValGlyLeuLeu 43
Db 22 GTGGCTTCAGGATCGCATCTGCAGACCTTTTCACGCGGCATAATGCGCTTCATGCTC 81
Qy 44 -----AspThrIleGlyIleAlaSerIle 51
Db 82 GGGCCTTCGTCTACGAAATTTTCGTAACTTCAGATCGTCAATGTCCAGCGCCATCGTTG 141
Qy 52 CysAlaLeuValValSerThrAlaProGluValMethHisAspThrArgPheVal 71
Db 142 TGTTCAGCGTTA-----GCAATGGCAGATTCAGAACCTTTCTG 180
Qy 72 ProThrLeuValGlyPheAlaValLeuGlyAlaSerPheTyrlsThrArgSerIle 91
Db 181 ACCAGTACAGCCGCTTCTTGTGGTGTAGGTCAAAATATATCCAGAGCCTCGCACACTTTC 240
Qy 92 IleProThrLeuLeuSerAlaLeu-----AlaTyrlsGlyLeuAlaTlPlys 106
Db 241 TTACCGGATCAGGTGAGTCAAGCAGCAGGCGAACCTTCTGAGCAGAGAACGAGCATGGCGA 300
Qy 107 ValMetAlaIleIle 111
Db 301 TGTTCAGCGATAGT 315
```

```
RESULT 5
US-09-711-164-134/c
; Sequence 134, Application US/09/711.164
; Patent No. 6589738
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERAPY
; FILE REFERENCE: ELITRA.008A
; CURRENT APPLICATION NUMBER: US/09/711.164
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/164415
; PRIOR FILING DATE: 1999-11-9
; NUMBER OF SEQ ID NOS: 469
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 134
; LENGTH: 333
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(333)
US-09-711-164-134
```

Alignment Scores:

```
Pred. No.: 0.317 Length: 333
Score: 77.00 Matches: 24
Percent Similarity: 41.9% Conservative: 20
Best Local Similarity: 22.9% Mismatches: 37
Query Match: 13.9% Indels: 24
DB: 3 Gaps: 3
```

US-10-073-293A-6 (1-111) x US-09-711-164-134 (1-333)

```
Qy 24 LeuArgLeuValGlyAsnAlaArgProThrLysArgGlyAlaValGlyLeuLeu 43
Db 299 GTGGCTTCAGGATCGCATCTGCAGACCTTTTCACGCGGCATAATGCGCTTCATGCTC 240
Qy 44 -----AspThrIleGlyIleAlaSerIle 51
Db 239 GGGCCTTCGTCTACGAAATTTTCGTAACTTCAGATCGTCAATGTCCAGCGCCATCGTTG 180
Qy 52 CysAlaLeuValValSerThrAlaProGluValMethHisAspThrArgPheVal 71
Db 179 TGTTCAGCGTTA-----GCAATGGCAGATTCAGAACCTTTCTG 141
Qy 72 ProThrLeuValGlyPheAlaValLeuGlyAlaSerPheTyrlsThrArgSerIle 91
Db 140 ACCAGTACAGCCGCTTCTTGTGGTGTAGGTCAAAATATATCCAGAGCCTCGCACACTTTC 81
Qy 92 IleProThrLeuLeuSerAlaLeu-----AlaTyrlsGlyLeuAlaTlPlys 106
Db 80 TTACCGGATCAGGTGAGTCAAGCAGCAGGCGAACCTTCTGAGCAGAGAACGAGCATGGCGA 21
Qy 107 ValMetAlaIleIle 111
Db 20 TGTTCAGCGATAGT 6
```

RESULT 6

```
US-09-492-709A-227/c
; Sequence 227, Application US/09/492.709A
; Patent No. 6720139
; GENERAL INFORMATION:
; APPLICANT: Zyskind, Judith
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Trawick, John
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Froelich, Jamie M.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; FILE REFERENCE: ELITRA.001A
; CURRENT APPLICATION NUMBER: US/09/492.709A
; CURRENT FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 227
; LENGTH: 333
; TYPE: DNA
; ORGANISM: E. Coli
US-09-492-709A-227
```

```
Alignment Scores:
Pred. No.: 0.317 Length: 333
Score: 77.00 Matches: 24
Percent Similarity: 41.9% Conservative: 20
Best Local Similarity: 22.9% Mismatches: 37
Query Match: 13.9% Indels: 24
DB: 3 Gaps: 3
```

US-10-073-293A-6 (1-111) x US-09-492-709A-227 (1-333)

```
Qy 24 LeuArgLeuValGlyAsnAlaArgProThrLysArgGlyAlaValGlyLeuLeu 43
Db 299 GTGGCTTCAGGATCGCATCTGCAGACCTTTTCACGCGGCATAATGCGCTTCATGCTC 240
```

```
Qy 44 -----AspThrIleGlyIleAlaSerIle 51
Db 239 GGGCTTCGCTACGAAATTTTCGTAACCTTCAGATCGTCAATGTACGCCCATCGTTG 180
Qy 52 CysAlaLeuLeuValSerThrAlaProGluValMetHisAspThrArgArgPheVal 71
Db 179 TGTTCAGCGTTA-----GCAATGGCAGATTCAGAACTTCTCTG 141
Qy 72 ProThrLeuValGlyPheAlaValLeuGlyAlaSerPheTyrLysThrArgSerIleIle 91
Db 140 ACCAGTACAGCGCTTCTTCTGTGTAGTCAAAATATCCAGAGCGCTCGCACACTTTC 81
Qy 92 IleProThrLeuLeuSerAlaLeu-----AlaTyrGlyLeuAlaTyrLys 106
Db 80 TTACCGGAATCAGCTCAGCAACAGGCGAACCTTCTGAGCAGAGAACGAGCATGGCGA 21
Qy 107 ValMetAlaIleIle 111
Db 20 TGTTTAGCGATAGT 6

RESULT 7
US-09-134-001C-2694
; Sequence 2694, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: GTC-007
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2694
; LENGTH: 1968
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2694

Alignment Scores:
Pred. No.: 5.23 Length: 1968
Score: 76.00 Matches: 34
Percent Similarity: 34.0% Conservative: 21
Best Local Similarity: 21.0% Mismatches: 39
Query Match: 13.8% Indels: 68
DB: 3 Gaps: 7

US-10-073-293A-6 (1-111) x US-09-134-001C-2694 (1-1968)
Qy 5 ValLeuLeuLeuGlyLeuLeuValGlyVal----- 14
Db 1435 ATTATGTTACTTGGTCTGTGATTTGGAGCTATGATGGCAATGTGACATGGCGGTCCTATT 1494
Qy 15 -----AlaAsnTyrCysPheArgTyrLeuProLeuArgLeuArgValGlyAsnAlaArg 32
Db 1495 AATAAAGCAGCTTATGATTT-----GCTACTGCTGCATTAACGGAAGAAATGCAGCT 1548
Qy 33 ProThrLysArgGlyAlaValGlyIleLeuLeuAspThrIleGlyIleAlaSerIleCys 52
Db 1549 CCTATTACTGCGGATGATTTGGTGTATGATCTCTCCATTAGCTATCGCTACA----- 1602
Qy 53 AlaLeuLeuValVal----- 57
Db 1603 GCAATGTTAACTTTAGAGAAATTCCTAAAGAGCAAGAGGCTCAATTTGTACCAAC 1662
Qy 58 -----SerThrAlaProGluValMetHisAspThr 67
Db 1663 TATGTCATGGGATTTATCATTTTATTACTGAAGGTGCCATCCCATTCGCAGCTGCCGATCCA 1722
Qy 68 ArgArgPheValProThr---LeuValGly----- 76
```

```
Db 1723 TTACGTGTGATTCCTCAATGATGTTGGCTCAGGTGTGGTGGAGCAATGCTTTAGGT 1782
Qy 77 -----PheAlaValLeuGlyAlaSerPhe 84
Db 1783 TTAGGATCAAGTATCAAAAGCACCTCATGGTGGAAATTTTCGTCAATTATTGGTACTGATTTT 1842
Qy 85 -----TyrLysThrArgSerIleIleIleProThrLeuLeuSerAlaLeuAla 100
Db 1843 AACCATATATCTCAAACTCTTATTTAGTTTGGGTAGCTTAGTTTTCAGCTTTAATT 1902
Qy 101 TyrGly 102
Db 1903 TATGTT 1908

RESULT 8
US-09-710-279-3573/c
; Sequence 3573, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 3573
; LENGTH: 4041
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-09-710-279-3573

Alignment Scores:
Pred. No.: 14.4 Length: 4041
Score: 76.00 Matches: 34
Percent Similarity: 34.0% Conservative: 21
Best Local Similarity: 21.0% Mismatches: 39
Query Match: 13.8% Indels: 68
DB: 3 Gaps: 7

US-10-073-293A-6 (1-111) x US-09-710-279-3573 (1-4041)
Qy 5 ValLeuLeuLeuGlyLeuLeuValGlyVal----- 14
Db 958 ATTATGTTACTTGGTCTGTGATTTGGAGCTATGATGGCAATGTGACATGGCGGTCCTATT 899
Qy 15 -----AlaAsnTyrCysPheArgTyrLeuProLeuArgLeuArgValGlyAsnAlaArg 32
Db 898 AATAAAGCAGCTTATGATTT-----GCTACTGCTGCATTAACGGAAGAAATGCAGCT 845
Qy 33 ProThrLysArgGlyAlaValGlyIleLeuLeuAspThrIleGlyIleAlaSerIleCys 52
Db 844 CCTATTACTGCGGATGATTTGGTGTATGATCTCTCCATTAGCTATCGCTACA----- 791
Qy 53 AlaLeuLeuValVal----- 57
Db 790 GCAATGTTAACTTTTAGAGAAATTCCTAAAGAGCAAGAGGCTCAATTTGTACCAAC 731
Qy 58 -----SerThrAlaProGluValMetHisAspThr 67
Db 730 TATGTCATGGGATTTATCATTTTATTACTGAAGGTGCCATCCCATTCGCAGCTGCCGATCCA 671
Qy 68 ArgArgPheValProThr---LeuValGly----- 76
Db 670 TTACGTGTGATTCCTCAATGATGTTGGCTCAGGTGTGGCTGAGCAATTTGCTTTAGGT 611
Qy 77 -----PheAlaValLeuGlyAlaSerPhe 84
```

```
Db 610 TTAGGATCAAGATATCAAAAGCACCTCATGGTGGAAATTTTCGTCAATATTGGTACTGATTTT 551
Qy 85 -----TyrLysThrArgSerIleIleProThrLeuLeuSerAlaLeuAla 100
Db 550 AACCATATACTTCAAACTCTTATTGCTTTAGTTGGGTACGTTAGTTTTCAGCTTTTAATT 491
Qy 101 TyrGly 102
Db 490 TATGGT 485

RESULT 9
US-09-710-279-3989
; Sequence 3989, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: P03480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 3989
; LENGTH: 4342
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-09-710-279-3989

Alignment Scores:
Pred. No.: 15.9 Length: 4342
Score: 76.00 Matches: 34
Percent Similarity: 34.0% Conservative: 21
Best Local Similarity: 21.0% Mismatches: 39
Query Match: 13.8% Indels: 68
DB: 3 Gaps: 7

US-10-073-293A-6 (1-111) x US-09-710-279-3989 (1-4342)

Qy 5 ValLeuLeuGlyLeuValGlyVal-----14
Db 172 ATTATGTTACTTGCTGTGTGATGGAGCTATGATGCAATTCACATGGCGGTCATTT 231
Qy 15 -----AlaAsnTyrCysPheArgTyrLeuProLeuArgLeuArgValGlyAsnAlaArg 32
Db 232 AATAAAGCAGCTTATGTATTT-----GCTACTGCTCATTAACGGAAGAAATGCAGCT 285
Qy 33 ProThrLysArgGlyAlaValGlyIleLeuLeuAspThrIleGlyIleAlaSerIleCys 52
Db 286 CCTATTACTGCAGCGATGATTTGGTGTGATGATTCCTCCATTAGCTATCGCTACA-----339
Qy 53 AlaLeuLeuValVal-----57
Db 340 GCAATGTTAATCTTTAGAGAAATTCACATAAGACGAAAGAGCTCAATTGTACCAAC 399
Qy 58 -----SerThrAlaProGluValMethHisAspThr 67
Db 400 TATGTCATGGATATCATTTATTACTGAAGGTGCCATCCATTCGCAGCTGCCATCCA 459
Qy 68 ArgArgPheValProThr-----LeuValGly-----76
Db 460 TTACGTGTGATTCCTCAATGATGGTTGGCTCAGGTGTGGCTGGAGCAATTGCTTAGGT 519
Qy 77 -----PheAlaValLeuGlyAlaSerPhe 84
Db 520 TTAGGATCAAGATATCAAAAGCACCTCATGGTGGAAATTTTCGTCAATATTGGTACTGATTTT 579
Qy 85 -----TyrLysThrArgSerIleIleProThrLeuLeuSerAlaLeuAla 100
```

```
Db 580 AACCATATACTTCAAACTCTTATTGCTTTAGTTGGGTACGTTAGTTTTCAGCTTTTAATT 639
Qy 101 TyrGly 102
Db 640 TATGGT 645

RESULT 10
US-09-270-767-14385/c
; Sequence 14385, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 14385
; LENGTH: 3209
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-14385

Alignment Scores:
Pred. No.: 14.1 Length: 3209
Score: 75.00 Matches: 35
Percent Similarity: 44.0% Conservative: 16
Best Local Similarity: 30.2% Mismatches: 33
Query Match: 13.6% Indels: 32
DB: 3 Gaps: 6

US-10-073-293A-6 (1-111) x US-09-270-767-14385 (1-3209)

Qy 5 ValLeuLeuGlyLeuValGlyValAlaAsnTyrCysPheArgTyrLeuProLeu 24
Db 804 GTTCTGATACTGTTCTCTCATCGTC-----AGAGGACTCGGTCTC 765
Qy 25 ArgLeuArgValGlyAsnAlaArgProThrLysArgGlyAlaValGlyIleLeuLeuAsp 44
Db 765 CGATTTCAGAGTC-----CGAGTCCCAATCAATGGAGTCATCGGAATCGTCATC---718
Qy 45 ThrIleGlyIleAlaSerIleCys-----52
Db 717 ---ATCGGGGGGCGCAGCTTTGCCGGCGGAGCTTCACAGCTTTCGGAACCGCTTCTGT 661
Qy 53 -----AlaLeuLeuValValSerThrAlaProGluValMethHisAspThrArg 68
Db 660 AGGCTTCACTCCCGCGCTGCTGTATCATCCAGCATCGCCCGCTCTCCAGGGCAAC 601
Qy 69 ArgPheValProThrLeuValGlyPheAlaValLeuGlyAlaSerPheTyrLysThrArg 88
Db 600 CACCTCGTC-----CTCGGCCCTCGCTCTCTCTGTCGGGTGCTTCACGGAACCGCA 550
Qy 89 SerIleIleIleProThrLeuLeuSerAlaLeuAlaTyrGlyLeuAla 104
Db 549 CAGATCGTCTC---CAAGTCTTGTATGTACTTTCGTACCTTCTGGCG 505

RESULT 11
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
```

```

; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:
Pred. No.: 3.9e+05 Length: 4403765
Score: 74.50 Matches: 24
Percent Similarity: 48.3% Conservative: 19
Best Local Similarity: 27.0% Mismatches: 41
Query Match: 13.5% Indels: 6
DB: 3 Gaps: 1

US-10-073-293A-6 (1-111) x US-09-103-840A-2 (1-4403765)

Qy 23 ProLeuArgLeuArgValGlyAsnAlaArgProThrLysArgGlyAlaValGlyLeu 42
Db 1250937 CCCTTGATGCTACTATTATGGCAACGCCACGCGCCACTGTGGCGCCACAGGGGGCGCCT 1250996

Qy 43 LeuAspThrIleGlyIleAlaSerIleCysAlaLeuLeuValValSerThrAlaProGlu 62
Db 1250997 TGGTCAGCGGTCGAGCTACTAGCTTGTCTGGTGTGGCCCTTACCCTGCTCGGAGA 1251056

Qy 63 ValMethHisAspThrArgArgPheValProThrLeuValGlyPheAlaValLeuGlyAla 82
Db 1251057 -----CAACGCCGCTGCCGTGATGGTGGCGCGGTGGACAGCGCA 1251101

Qy 83 SerPheTyrLysThrArgSerIleIleIleProThrLeuLeuSerAlaLeuAlaTyrGly 102
Db 1251102 CCGGCCCAACGGCTGGTTCGACCGG-CTCCCCCGCCTAACGCTACGGGTGCCTTCGTC 1251160

Qy 103 LeuAlaTrpLysValMetAlaIleIle 111
Db 1251161 GTCTGCCAGGAGCTTTCCCGGTGATG 1251187

RESULT 12
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Alignment Scores:
Pred. No.: 3.91e+05 Length: 4411529
Score: 74.50 Matches: 24
Percent Similarity: 48.3% Conservative: 19
Best Local Similarity: 27.0% Mismatches: 41
Query Match: 13.5% Indels: 6
DB: 3 Gaps: 1

```

```

US-10-073-293A-6 (1-111) x US-09-103-840A-1 (1-4411529)

Qy 23 ProLeuArgLeuArgValGlyAsnAlaArgProThrLysArgGlyAlaValGlyLeu 42
Db 1251409 CCCTTGATGCTACTATTATGGCAACGCCACGCGCCACTGTGGCGCCACAGGGGGCGCCT 1251468

Qy 43 LeuAspThrIleGlyIleAlaSerIleCysAlaLeuLeuValValSerThrAlaProGlu 62
Db 1251469 TGGTCAGCGGTCGAGCTACTAGCTTGTCTGGTGTGGCCCTTACCCTGCTCGGAGA 1251528

Qy 63 ValMethHisAspThrArgArgPheValProThrLeuValGlyPheAlaValLeuGlyAla 82
Db 1251529 -----CAACGCCGCTGCCGTGATGGTGGCGCGGTGGACAGCGCA 1251573

Qy 83 SerPheTyrLysThrArgSerIleIleIleProThrLeuLeuSerAlaLeuAlaTyrGly 102
Db 1251574 CCGGCCCAACGGCTGGTTCGACCGG-CTCCCCCGCCTAACGCTACGGGTGCCTTCGTC 1251632

Qy 103 LeuAlaTrpLysValMetAlaIleIle 111
Db 1251633 GTCTGCCAGGAGCTTTCCCGGTGATG 1251659

RESULT 13
US-09-252-991A-6623/c
; Sequence 6623, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6623
; LENGTH: 708
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6623

Alignment Scores:
Pred. No.: 2.29 Length: 708
Score: 74.00 Matches: 40
Percent Similarity: 39.0% Conservative: 6
Best Local Similarity: 33.9% Mismatches: 41
Query Match: 13.4% Indels: 31
DB: 3 Gaps: 6

US-10-073-293A-6 (1-111) x US-09-252-991A-6623 (1-708)

Qy 20 ArgTyrTrpProLeuArgLeuArg-ValGlyAsnAla----- 31
Db 568 CGATACCTTCCTCGCCGCTACCGCGCGCGCGCGGTACCGCGCGCGGTCTTCACCTT 509

Qy 32 -----ArgProThrLysArgGlyAlaValGlyLeuLe 43
Db 508 CGCGCGCTTCCTCGCGCGAGCTCGGGCGCGCGCGCGCGAGCGAGTGGAGCGCGCTGCT 449

Qy 43 uAspThrIleGlyIleAlaSerIleCysAlaLeuLeuValValSerThrAlaPro----- 61
Db 448 GTGCCTGCTGCGGATCTTCCTCGCGCTGCTTCCTGCTGCTGCGCGGTGCTTCATCTGT 389

Qy 62 -GluValMetHisAspThrArgArgPheValProThrLeuValGlyPhe-----AlaVa 79
Db 388 GGAACGCTCGCGCACCAACCGCGGTACCGGGACACGCTGGCGCGGTCAACGCGGGGT 329

Qy 79 lLeuGly-----AlaSerPheTyrLys-----ThrArgSerIleIlePr 93
Db 328 GGTGCGGCTGCTGCTGCGCGGCTGTACCGCGCTCTGGACCGAGCATCTCTGCGGCC 269

```


Percent Similarity: 44.4% Conservative: 18
 Best Local Similarity: 27.8% Mismatches: 38
 Query Match: 13.4% Indels: 22
 DB: 3 Gaps: 5

US-10-073-293A-6 (1-111) x US-09-603-208A-229 (1-729)

```

QY      5 ValLeuLeuLeuGlyLeuValGlyValAlaAlaSerThrCysPheArgTyrLeuProLeu 24
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 278 ATCATGCTCATCGGCTGGTATTGGTCTGGCCATGGATTACAGATCTTCTCGTTACT 337
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      25 ArgLeuArg-----ValGlyAsnAlaArgProThr----- 34
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 338 CGTATGCGTGAAGGCTTCACCAAGGGCAAGACTGGGGCAACGCAACGTCGAATGGTTTC 397
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      35 LysArgGlyAlaValGlyIleLeuLeuAspThrIleGlyIleAlaSerIleCysAlaLeu 54
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 398 AAGCAGCGTGCC-----CGCGTGTGTCACTGTCTGCGCGCGCTG 433
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      55 LeuValValSerThrAlaProGluValMethHisAspThrArgArgPheValProThrLeu 74
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 434 ATCATGGTGTCTGTTCGGCGCATTCATAGCGGAGACATGGCGTTTATTAGACCATG 493
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      75 ValGlyPheAlaValLeuGlyAlaSerPheTyrLys-----ThrArgSerIleIle 91
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 494 ---GGCTTTGCTCTGGCGGTTGTGTCTTCGATGCCTTCGTTGTTCGATGATGATT 550
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      92 IleProThrLeuLeuSerAlaLeu 99
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 551 ATCCCTGCACAATGTTCTCGCTT 574
    
```

Search completed: February 16, 2006, 19:10:52
 Job time : 381.334 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 16, 2006, 17:44:58 ; Search time 429.034 Seconds
(without alignments)
2139.461 Million cell updates/sec

Title: US-10-073-293A-6
Perfect score: 552
Sequence: 1 MSTEVLLGLLVGVNYCFR.....IPTLLSALAYGLAWKMAII 111

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/abs/ABSSWEB_spool/US10073293/runat_15022006_120740_1106/app_query.fasta_1
-DB=Published Applications NA_Main -QMT=fascap -SUFFIX=p2n.rnpbm
-MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1
-MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct
-THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext
-HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abes03p
-USER=US10073293 @CGN 1_1_2064 @runat_15022006_120740_1106 -NCPU=6 -ICPU=3
-NO_MMAP -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA_Main:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	552	100.0	336	9	US-10-073-293A-5
2	98	17.8	172	6	US-10-360-201-18
3	94	17.0	670	7	US-10-424-599-80195
c 4	80.5	14.6	507	8	US-10-425-115-112500
5	79.5	14.4	3804	7	US-10-424-599-18810
6	78	14.1	1593	7	US-10-282-122A-27599
c 7	78	14.1	1879	3	US-09-728-260-1

c 8	78	14.1	1879	3	US-09-798-412-1	Sequence 1, Appli
c 9	78	14.1	1879	5	US-10-325-917-1	Sequence 1, Appli
c 10	77.5	14.0	1326	7	US-10-282-122A-26760	Sequence 26760, A
11	77.5	14.0	19988	7	US-10-672-787-10	Sequence 10, Appl
12	77	13.9	330	6	US-10-287-274-57	Sequence 57, Appl
c 13	77	13.9	333	3	US-09-741-669-127	Sequence 127, App
c 14	77	13.9	333	3	US-09-912-020-227	Sequence 227, App
c 15	77	13.9	333	3	US-09-815-242-6257	Sequence 6257, Ap
c 16	77	13.9	333	6	US-10-287-274-134	Sequence 134, App
c 17	77	13.9	333	7	US-10-282-122A-1248	Sequence 1248, Ap
c 18	77	13.9	333	7	US-10-282-122A-6478	Sequence 6478, Ap
c 19	77	13.9	333	8	US-10-771-241-227	Sequence 227, App
c 20	77	13.9	654	7	US-10-437-963-70639	Sequence 70639, A
c 21	77	13.9	924	6	US-10-369-493-27216	Sequence 27216, A
c 22	77	13.9	1107	7	US-10-437-963-12803	Sequence 12803, A
c 23	77	13.9	10944	9	US-10-893-671-40	Sequence 40, Appl
c 24	76	13.8	330	7	US-10-282-122A-37364	Sequence 37364, A
c 25	76	13.8	333	3	US-09-815-242-9872	Sequence 9872, Ap
c 26	76	13.8	333	7	US-10-282-122A-7550	Sequence 7550, Ap
c 27	76	13.8	333	7	US-10-282-122A-39532	Sequence 39532, A
c 28	76	13.8	405	7	US-10-282-122A-37415	Sequence 37415, A
c 29	76	13.8	635	8	US-10-425-115-157416	Sequence 157416,
c 30	76	13.8	1544	7	US-10-092-900A-313	Sequence 313, App
c 31	76	13.8	1617	7	US-10-437-963-5392	Sequence 5392, Ap
c 32	76	13.8	1950	7	US-10-282-122A-34484	Sequence 34484, A
c 33	76	13.8	1968	7	US-10-724-972A-3483	Sequence 3483, Ap
c 34	75.5	13.7	826	8	US-10-425-115-106830	Sequence 106830,
c 35	75.5	13.7	999	9	US-10-501-282-5129	Sequence 5129, Ap
c 36	75.5	13.7	999	9	US-10-501-282-5131	Sequence 5131, Ap
c 37	75.5	13.7	999	9	US-10-501-282-5133	Sequence 5133, Ap
c 38	75.5	13.7	999	9	US-10-501-282-5135	Sequence 5135, Ap
c 39	75.5	13.7	999	9	US-10-501-282-5137	Sequence 5137, Ap
c 40	75.5	13.7	999	9	US-10-501-282-5139	Sequence 5139, Ap
c 41	75.5	13.7	999	9	US-10-501-282-5141	Sequence 5141, Ap
c 42	75.5	13.7	1754382	9	US-10-501-282-6651	Sequence 6651, Ap
c 43	75	13.6	3290	10	US-11-097-143-10172	Sequence 10172, A
c 44	75	13.6	5567	10	US-11-097-143-10171	Sequence 10171, A
c 45	74.5	13.5	1371	6	US-10-369-493-45062	Sequence 45062, A

ALIGNMENTS

RESULT 1

US-10-073-293A-5
; Sequence 5, Application US/10073293A
; Publication No. US20050239175A1
; GENERAL INFORMATION:
; APPLICANT: TABOLINA, EKATERINA
; APPLICANT: RYBAK, KONSTANTIN
; APPLICANT: KHOURGES, EVGENI
; APPLICANT: VOROSHILOVA, ELVIRA
; APPLICANT: GUSYATINER, MIKHAIL
; TITLE OF INVENTION: METHOD FOR PRODUCING L-AMINO ACID USING BACTERIA BELONGING TO THE
; FILE REFERENCE: 219594USO
; CURRENT APPLICATION NUMBER: US/10/073,293A
; CURRENT FILING DATE: 2002-02-13
; PRIOR APPLICATION NUMBER: RU 2001103865
; PRIOR FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: RU 2001104998
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: RU 2001104999
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: RU 2001117632
; PRIOR FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: RU 2001117633
; PRIOR FILING DATE: 2001-06-28
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 336
; TYPE: DNA
; ORGANISM: Escherichia coli


```

; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27599
; LENGTH: 1593
; TYPE: DNA
; ORGANISM: Mycobacterium leprae
US-10-282-122A-27599

Alignment Scores:
Pred. No.: 4.62 Length: 1593
Score: 78.00 Matches: 31
Percent Similarity: 40.5% Conservative: 14
Best Local Similarity: 27.9% Mismatches: 32
Query Match: 14.1% Indels: 34
DB: 7 Gaps: 5

US-10-073-293A-6 (1-111) x US-10-282-122A-27599 (1-1593)
Qy 4 GluValLeu-----LeuLeuGlyLeuLeuValGlyValAlaAsnTyrCysPheArg 20
Db 166 GAGGTGCTGTCGGGTGATCTGTTGCTGCATTGGCTGCTGCATCCACACAGCGCGCTCGG 225
Qy 21 TyrLeuProLeuArgLeuArgValGlyAsnAlaArgProThrLysArgGlyAlaValGly 40
Db 226 TATGCCGCGTGTGCGCTC-----GAACGAGGTGCTGTCGCG 261
Qy 41 IleLeuLeuAspThrIleGlyLeuAlaSerIleCysAlaLeuLeuValValSerThrAla 60
Db 262 GTGCTTACCGACGTGCGAGGGTCCCGAGCTGACAGCC-----CAGGCCAGAGCGGTG 315
Qy 61 ProGluValMetHisAspThrArgArgPheValProThrLeuValGlyPheAlaValLeu 80
Db 316 CCGTACTCATACATCCGGAACCCCGC-----AGCGTGCTC 351
Qy 81 GlyAlaSerPheTyrLysThrArgSerIleIleIleProThrLeuLeuSerAlaLeuAla 100
Db 352 GCGCGT-----CTTGCGCGCGCGGCA 372
Qy 101 TyrGlyLeuAlaTrpLysValMetAlaIleIle 111
Db 373 TATGACATCCTCCACCGGATGACGGTGTGT 405

RESULT 7
US-09-728-260-1/c
; Sequence 1, Application US/09728260
; Publication No. US20020086980A1
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
; TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
; FILE REFERENCE: 07334-319001
; CURRENT APPLICATION NUMBER: US/09/728,260
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: US 09/685,791
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: US 09/513,904
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: US 09/507,533
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: US 60/168,780
; PRIOR FILING DATE: 1999-12-03
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1879
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (113) ... (1720)
US-09-728-412-1

Alignment Scores:
Pred. No.: 5.84 Length: 1879
Score: 78.00 Matches: 34
Percent Similarity: 46.4% Conservative: 17
Best Local Similarity: 30.9% Mismatches: 34
Query Match: 14.1% Indels: 25
DB: 3 Gaps: 5

US-10-073-293A-6 (1-111) x US-09-728-260-1 (1-1879)
Qy 2 SerTyrGluValLeuLeuLeuGlyLeuLeuVal-----GlyValAlaAsnTyrCysPhe 19
Db 1850 AGCTCAGAGGTCAATTTATTAGGCTGCATTTTTTTCTTAATCTGTGTACACAGTATGTGTA 1791
Qy 20 ArgTyr-----LeuProLeuArgLeuArgValGlyAsnAla-ArgPr 33
Db 1790 GCGTACAGTAAAAAACCAGCAGCATCCCTTCACATTCACAGACGCTCAGCCTCGCG 1731
Qy 33 oThrLysArgGlyAlaValGlyIleLeuLeuAspThrIleGlyIleAlaSerIleCysAl 53
Db 1730 TGGTCTCTAGGAGCCCTCGT-----GTCGGTGTGTGCTGCTGCTGCG 1686
Qy 53 aLeuLeuValVal-----SerThrAlaProGluValMetHisAspThrArgArgPheVa 71
Db 1685 TATTCCCGTATCCCTTCTCCCTGTCGCCAGCTGTTCTGCATCTTGGGAGGCCCGCT 1626
Qy 71 lProThrLeuValGlyPheAlaValLeuLeuGlyAlaSerPheTyrLysThrArgSerIleI 91
Db 1625 TCCTGCGGTAGTCTCGAAGCTCTCCTTGAGGCGCG-----1589
Qy 91 elleProThrLeuLeuSerAlaLeuAla 100
Db 1588 ----CCGCTCCTTCTCGCGGGGCTCGCT 1565

RESULT 8
US-09-798-412-1/c
; Sequence 1, Application US/09798412
; Publication No. US20030109428A1
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
; TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
; FILE REFERENCE: 07334-327001
; CURRENT APPLICATION NUMBER: US/09/798,412
; CURRENT FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: US 09/728,260
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: US 09/685,791
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: US 09/513,904
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: US 09/507,533
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: US 60/168,780
; PRIOR FILING DATE: 1999-12-03
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1879
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (113) ... (1720)
US-09-798-412-1

Alignment Scores:
Pred. No.: 5.84 Length: 1879
Score: 78.00 Matches: 34

```

Percent Similarity: 46.4% Conservative: 17
 Best Local Similarity: 30.9% Mismatches: 34
 Query Match: 14.1% Indels: 25
 DB: 3 Gaps: 5

US-10-073-293A-6 (1-111) x US-09-798-412-1 (1-1879)

Qy 2 SerTyrGluValLeuLeuLeuGlyLeuVal-----GlyValAlaAseTyrCysPhe 19
 Db 1850 AGCTCAGAGGTCAATTTATTAGGCTGATTTTCTAATACTTGTACACAGGTATGTGA 1791
 Qy 20 ArgTyr-----LeuProLeuArgLeuValGlyAseAla-ArgPr 33
 Db 1790 GCGTACAGTAAAAAACCCGACGATCCCTTCACATTTACACAGCCTCAGCCTCGCG 1731
 Qy 33 oThrLysArgGlyAlaValGlyLeuLeuAspThrIleGlyLeuAseRlleCysAl 53
 Db 1730 TGGTCTCTAGAGCCCTCGCT-----GTCGGTGTTCGCTGCTGCTCG 1686
 Qy 53 aLeuLeuValVal-----SerThrAlaProGluValMetHisAspThrArgPheVa 71
 Db 1685 TATTCCCGTGATCCCTCTCTCCCTGTCGCCAGCTGTTCTGCATCTTCGGGAGCGCCGCT 1626
 Qy 71 lProThrLeuValGlyPheAlaValLeuGlyAlaSerPheTyrLysThrArgSerlleIl 91
 Db 1625 TCCTGCGGTAGTTCTCGAAGCTCTCTTCGAGGCGCG-----1589

Qy 91 elleProThrLeuLeuSerAlaLeuAla 100
 Db 1588 -----CCGCTCTCTTCGCGGGGCTCGCT 1565

RESULT 9
 US-10-325-917-1/c
 ; Sequence 1, Application US/10325917
 ; Publication No. US20030113787A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bertin, John
 ; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
 ; FILE REFERENCE: 07334-327001
 ; CURRENT APPLICATION NUMBER: US/10/325,917
 ; CURRENT FILING DATE: 2002-12-20
 ; PRIOR APPLICATION NUMBER: US/09/798,412
 ; PRIOR FILING DATE: 2001-03-02
 ; PRIOR APPLICATION NUMBER: US 09/728,260
 ; PRIOR FILING DATE: 2000-12-01
 ; PRIOR APPLICATION NUMBER: US 09/685,791
 ; PRIOR FILING DATE: 2000-10-10
 ; PRIOR APPLICATION NUMBER: US 09/513,904
 ; PRIOR FILING DATE: 2000-02-25
 ; PRIOR APPLICATION NUMBER: US 09/507,533
 ; PRIOR FILING DATE: 2000-02-18
 ; PRIOR APPLICATION NUMBER: US 60/168,780
 ; PRIOR FILING DATE: 1999-12-03
 ; NUMBER OF SEQ ID NOS: 19
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 1
 ; LENGTH: 1879
 ; TYPE: DNA
 ; ORGANISM: Rattus norvegicus
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (113)...(1720)
 US-10-325-917-1

Alignment Scores:
 Pred. No.: 5.84 Length: 1879
 Score: 78.00 Matches: 34
 Percent Similarity: 46.4% Conservative: 17
 Best Local Similarity: 30.9% Mismatches: 34
 Query Match: 14.1% Indels: 25
 DB: 5 Gaps: 5

US-10-073-293A-6 (1-111) x US-10-325-917-1 (1-1879)

Qy 2 SerTyrGluValLeuLeuLeuGlyLeuVal-----GlyValAlaAseTyrCysPhe 19
 Db 1850 AGCTCAGAGGTCAATTTATTAGGCTGATTTTCTAATACTTGTACACAGGTATGTGA 1791
 Qy 20 ArgTyr-----LeuProLeuArgLeuValGlyAseAla-ArgPr 33
 Db 1790 GCGTACAGTAAAAAACCCGACGATCCCTTCACATTTACACAGCCTCAGCCTCGCG 1731
 Qy 33 oThrLysArgGlyAlaValGlyLeuLeuAspThrIleGlyLeuAseRlleCysAl 53
 Db 1730 TGGTCTCTAGAGCCCTCGCT-----GTCGGTGTTCGCTGCTGCTCG 1686
 Qy 53 aLeuLeuValVal-----SerThrAlaProGluValMetHisAspThrArgPheVa 71
 Db 1685 TATTCCCGTGATCCCTCTCTCCCTGTCGCCAGCTGTTCTGCATCTTCGGGAGCGCCGCT 1626
 Qy 71 lProThrLeuValGlyPheAlaValLeuGlyAlaSerPheTyrLysThrArgSerlleIl 91
 Db 1625 TCCTGCGGTAGTTCTCGAAGCTCTCTTCGAGGCGCG-----1589

Qy 91 elleProThrLeuLeuSerAlaLeuAla 100
 Db 1588 -----CCGCTCTCTTCGCGGGGCTCGCT 1565

RESULT 10
 US-10-282-122A-26760
 ; Sequence 26760, Application US/10282122A
 ; Publication No. US20040029129A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John
 ; APPLICANT: Carr, Grant
 ; APPLICANT: Yamamoto, Robert
 ; APPLICANT: Forsyth, R.
 ; APPLICANT: Xu, H.
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 ; FILE REFERENCE: ELITRA.034A
 ; CURRENT APPLICATION NUMBER: US/10/282,122A
 ; CURRENT FILING DATE: 2003-02-20
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/230,335
 ; PRIOR FILING DATE: 2000-09-06
 ; PRIOR APPLICATION NUMBER: 60/230,347
 ; PRIOR FILING DATE: 2000-09-09
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/267,636
 ; PRIOR FILING DATE: 2001-02-09
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 78614
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 26760
 ; LENGTH: 1326
 ; TYPE: DNA

```
; ORGANISM: Moraxella catarrhalis
US-10-282-122A-26760

Alignment Scores:
Pred. No.: 4.18 Length: 1326
Score: 77.50 Matches: 32
Percent Similarity: 40.3% Conservative: 20
Best Local Similarity: 24.8% Mismatches: 36
Query Match: 14.0% Indels: 41
DB: 7 Gaps: 6

US-10-073-293A-6 (1-111) x US-10-282-122A-26760 (1-1326)

Qy 9 GlyLeuLeuValGlyValAlaAsnTyrCysPheArgTyr----- 21
Db 763 GGTGGCTCATCGGCTTAATCAGTCTTGCACCGCTTTTACCACATTTTGCACCAATGAT 822
Qy 22 -----LeuProLeuArgLeuArgValGlyAsnAlaAArgProThrLys 35
Db 823 TTGATTTTAGATGCTGACTTACCATGCAATATTGGTGGCAATGTTTCATCCATGG--- 879
Qy 36 ArgGlyAlaValGlyLeuLeuAspThrIleGlyIleAlaSerIleCysAlaLeuLeu 55
Db 880 -----CTTGGCACTTTGATGAGTTTATCATTTTGGC-----ATG 915
Qy 56 ValValSerThrAlaProGluValMethis-----Asp 66
Db 916 ATTTTAAATACCGCCATCAGTTATATTACTCGTGGCCAAAGCGTTTTCAGGTGATGAC 975
Qy 67 ThrArgArgPheValProThrLeuValGlyPheAlaValLeuGly----- 81
Db 976 AACGAGCGCTTAAATGATATTGATGGCTGGTTGGTGGCTTTATCTTAAGCTTT 1035
Qy 82 AlaSerPheTyrLysThrArgSerIleIleIleProThr-----LeuLeu 96
Db 1036 GCAGGTTTTAAATAAACTGGTCTCAATCATGTTCCCATCATTTGGCTACATGGTATGTTG 1095
Qy 97 SerAlaLeuAlaTyrGlyLeuAlaTyr 105
Db 1096 CTTCTTGTGCTGTTGTTGTTGGCATGG 1122

RESULT 11
US-10-672-787-10
; Sequence 10, Application US/10672787
; Publication No. US20040067554A1
; GENERAL INFORMATION:
; APPLICANT: PATERSON, Chandra
; APPLICANT: BERG, Kim, L.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
; FILE REFERENCE: ELITRA.025C1
; CURRENT APPLICATION NUMBER: US/10/672,787
; PRIOR FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: 09/596,002
; PRIOR FILING DATE: 2000-06-16
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PERL Program
; SEQ ID NO 10
; LENGTH: 19988
; TYPE: DNA
; ORGANISM: Moraxella catarrhalis
US-10-672-787-10

Alignment Scores:
Pred. No.: 200 Length: 19988
Score: 77.50 Matches: 32
Percent Similarity: 40.3% Conservative: 20
Best Local Similarity: 24.8% Mismatches: 36
Query Match: 14.0% Indels: 41
DB: 7 Gaps: 6

US-10-073-293A-6 (1-111) x US-10-672-787-10 (1-19988)
```

```
Qy 9 GlyLeuLeuValGlyValAlaAsnTyrCysPheArgTyr----- 21
Db 2843 GGTGGCTCATCGGCTTAATCAGTCTTGCACCGCTTTTACCACATTTGCCAAACATGAT 2902
Qy 22 -----LeuProLeuArgLeuArgValGlyAsnAlaAArgProThrLys 35
Db 2903 TTGATTTTAGATGCTGACTTACCATGCAATATTGGTGGCAATGTTTCATCCATGG--- 2959
Qy 36 ArgGlyAlaValGlyLeuLeuAspThrIleGlyIleAlaSerIleCysAlaLeuLeu 55
Db 2960 -----CTTGGCACTTTGATGAGTTTATCATTTTGGC-----ATG 2995
Qy 56 ValValSerThrAlaProGluValMethis-----Asp 66
Db 2996 ATTTTAAATACCGCCATCAGTTATATTACTCGTGGCCAAAGCGTTTTCAGGTGATGAC 3055
Qy 67 ThrArgArgPheValProThrLeuValGlyPheAlaValLeuGly----- 81
Db 3056 AACGAGCGCTTAAATGATATTGATGGCTGGTTGGTGGCTTTATCTTAAGCTTT 3115
Qy 82 AlaSerPheTyrLysThrArgSerIleIleIleProThr-----LeuLeu 96
Db 3116 GCAGGTTTTAAATAAACTGGTCTCAATCATGTTCCCATCATTTGGCTACATGGTATGTTG 3175
Qy 97 SerAlaLeuAlaTyrGlyLeuAlaTyr 105
Db 3176 CTTCTTGTGCTGTTGTTGTTGGCATGG 3202

RESULT 12
US-10-287-274-57
; Sequence 57, Application US/10287274
; Publication No. US20030181408A1
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERETO
; FILE REFERENCE: ELITRA.008DV1
; CURRENT APPLICATION NUMBER: US/10/287,274
; CURRENT FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: US 60/164415
; PRIOR FILING DATE: 1999-11-09
; PRIOR APPLICATION NUMBER: US 09/711164
; PRIOR FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 469
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 57
; LENGTH: 330
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-287-274-57

Alignment Scores:
Pred. No.: 0.679 Length: 330
Score: 77.00 Matches: 24
Percent Similarity: 41.9% Conservative: 20
Best Local Similarity: 22.9% Mismatches: 37
Query Match: 13.9% Indels: 24
DB: 6 Gaps: 3

US-10-073-293A-6 (1-111) x US-10-287-274-57 (1-330)

Qy 24 LeuArgLeuArgValGlyAsnAlaAArgProThrLysArgGlyAlaValGlyLeuLeu 43
Db 22 GTGGCTTCAGGATCGCATCTGCACGACCTTTTGCACGCGCAATAATGCGCTTCATGCTC 81
Qy 44 -----AspThrIleGlyIleAlaSerIle 51
Db 82 GGGCTTCGCTCTACGAAATTTTCGTAACCTTTCAGATCGTCAATGTCAGCGCCATCGTTG 141
Qy 52 CysAlaLeuLeuValSerThrAlaProGluValMetHisAspThrArgArgPheVal 71
Db 142 TGTTCAGCGTTA-----GCAATGGCAGATTCAGAACTTTCTT 180
```


Qy 72 ProThrLeuValGlyPheAlaValLeuGlyAlaSerPheTyrLysThrArgSerIleIle 91
 Db 181 ACCAGTACAGCGCTTCTGTTGGTGTAGTCAAAATATCCAGAGCCTCGACACTTC 240
 Qy 92 IleProThrLeuLeuSerAlaLeu-----AlaTyrGlyLeuAlaTrpLys 106
 Db 241 TTACCGCAATCAGGTTCAGCAACAGGCGAACCTTCTGAGCAGAAGAACGAGCATGCGCA 300
 Qy 107 ValMetAlaIleIle 111
 Db 301 TGTTCAGCATAGTT 315

RESULT 13

US-09-741-669-127/c
 ; Sequence 127, Application US/09741669
 ; Patent No. US2002022718A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Forsyth, R. Allyn
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; TITLE OF INVENTION: Genes identified as required for
 ; proliferation of E. coli
 ; FILE REFERENCE: ELITRA.009A
 ; CURRENT APPLICATION NUMBER: US/09/741,669
 ; PRIOR FILING DATE: 2000-12-19
 ; PRIOR APPLICATION NUMBER: US 60/173005
 ; PRIOR FILING DATE: 1999-12-23
 ; NUMBER OF SEQ ID NOS: 481
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 127
 ; LENGTH: 333
 ; TYPE: DNA
 ; ORGANISM: Escherichia coli
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)...(333)
 ; US-09-741-669-127

Alignment Scores:

Pred. No.: 0.688 Length: 333
 Score: 77.00 Matches: 24
 Percent Similarity: 41.9% Conservative: 20
 Best Local Similarity: 22.9% Mismatches: 37
 Query Match: 13.9% Indels: 24
 DB: 3 Gaps: 3

US-10-073-293A-6 (1-111) x US-09-741-669-127 (1-333)

Qy 24 LeuArgLeuValGlyAlaValLeuGlyAlaSerPheTyrLysThrArgSerIleLeuLeu 43
 Db 299 GTGCGCTTCAGGATCGCATCTGCAGCACCTTTTGCACGCGCATATGCGCTTCATGCTC 240
 Qy 44 -----AspThrIleGlyIleAlaSerIle 51
 Db 239 GGGCCTTCGCTACGAAATTTTCGTAATTTTCAGATCGTCAATGTCAGCGCCATCGTTG 180
 Qy 52 CysAlaLeuLeuValSerThrAlaProGluValMethHisAspThrArgArgPheVal 71
 Db 179 TGTTCAGCGTTA-----GCAATGGCAGATCCAGAACCTTCTTCTG 141
 Qy 72 ProThrLeuValGlyPheAlaValLeuGlyAlaSerPheTyrLysThrArgSerIleIle 91
 Db 140 ACCAGTACAGCGCTTCTGTTGGTGTAGTCAAAATATCCAGAGCCTCGACACTTC 81
 Qy 92 IleProThrLeuLeuSerAlaLeu-----AlaTyrGlyLeuAlaTrpLys 106
 Db 80 TTACCGCAATCAGGTTCAGCAACAGGCGAACCTTCTGAGCAGAAGAACGAGCATGCGCA 21
 Qy 107 ValMetAlaIleIle 111
 Db 20 TGTTCAGCATAGTT 6

RESULT 14

US-09-912-020-227/c
 ; Sequence 227, Application US/09912020
 ; Patent No. US20020045592A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Trawick, John
 ; APPLICANT: Forsyth, R. Allyn
 ; APPLICANT: Froelich, Jamie M.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
 ; FILE REFERENCE: ELITRA.001DV1
 ; CURRENT APPLICATION NUMBER: US/09/912,020
 ; CURRENT FILING DATE: 2001-07-23
 ; PRIOR APPLICATION NUMBER: 09/492,709
 ; PRIOR FILING DATE: 2000-01-27
 ; PRIOR APPLICATION NUMBER: 60/117,405
 ; PRIOR FILING DATE: 1999-01-27
 ; NUMBER OF SEQ ID NOS: 485
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 227
 ; LENGTH: 333
 ; TYPE: DNA
 ; ORGANISM: E. Coli
 ; US-09-912-020-227

Alignment Scores:

Pred. No.: 0.688 Length: 333
 Score: 77.00 Matches: 24
 Percent Similarity: 41.9% Conservative: 20
 Best Local Similarity: 22.9% Mismatches: 37
 Query Match: 13.9% Indels: 24
 DB: 3 Gaps: 3

US-10-073-293A-6 (1-111) x US-09-912-020-227 (1-333)

Qy 24 LeuArgLeuValGlyAlaValLeuGlyAlaSerPheTyrLysThrArgSerIleLeuLeu 43
 Db 299 GTGCGCTTCAGGATCGCATCTGCAGCACCTTTTGCACGCGCATATGCGCTTCATGCTC 240
 Qy 44 -----AspThrIleGlyIleAlaSerIle 51
 Db 239 GGGCCTTCGCTACGAAATTTTCGTAATTTTCAGATCGTCAATGTCAGCGCCATCGTTG 180
 Qy 52 CysAlaLeuLeuValSerThrAlaProGluValMethHisAspThrArgArgPheVal 71
 Db 179 TGTTCAGCGTTA-----GCAATGGCAGATCCAGAACCTTCTTCTG 141
 Qy 72 ProThrLeuValGlyPheAlaValLeuGlyAlaSerPheTyrLysThrArgSerIleIle 91
 Db 140 ACCAGTACAGCGCTTCTGTTGGTGTAGTCAAAATATCCAGAGCCTCGACACTTC 81
 Qy 92 IleProThrLeuLeuSerAlaLeu-----AlaTyrGlyLeuAlaTrpLys 106
 Db 80 TTACCGCAATCAGGTTCAGCAACAGGCGAACCTTCTGAGCAGAAGAACGAGCATGCGCA 21
 Qy 107 ValMetAlaIleIle 111
 Db 20 TGTTCAGCATAGTT 6

RESULT 15

US-09-815-242-6257/c
 ; Sequence 6257, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel

```

; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6257
; LENGTH: 333
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(333)
US-09-815-242-6257

```

Alignment Scores:

Pred. No.:	0.688	Length:	333
Score:	77.00	Matches:	24
Percent Similarity:	41.9%	Conservative:	20
Best Local Similarity:	22.9%	Mismatches:	37
Query Match:	13.9%	Indels:	24
DB:	3	Gaps:	3

US-10-073-293A-6 (1-111) x US-09-815-242-6257 (1-333)

```

Qy 24 LeuArgLeuValGlyAsnAlaArgProThrLysArgGlyAlaValGlyLeuLeu 43
Db 299 GTGGCTTCAGGATCGCATCTGCAGACCTTTTGCACGGCGCATAAATCGGCTTCATGCTC 240
Qy 44 -----AspThrIleGlyLeuAlaSerIle 51
Db 239 GGGCTTCGCTACGAAATTTTCGTAACTTTTCAGATCGTCATGTCAGCGCCATCGTTG 180
Qy 52 CysAlaLeuLeuValValSerThrAlaProGluValMetHisAspThrArgArgPheVal 71
Db 179 TGTTCAGCGTTA-----GCAATGGCAGATTCAGAACTTTCTTG 141
Qy 72 ProThrLeuValGlyPheAlaValLeuGlyAlaSerPheTyrLysThrArgSerIleIle 91
Db 140 ACCAGTACAGCGCTTTCTTGTGTGTAGGTCAAAATATCCAGAGCTCGCACACTTTC 81
Qy 92 IleProThrLeuSerAlaLeu-----AlaTyrGlyLeuAlaTrpLys 106
Db 80 TTACCGCAATCAGTCTAGTCAACAGGCGAACCTTCTGAGCAGAGAACGAGCATGGCGA 21
Qy 107 ValMetAlaIleIle 111
Db 20 TGTTCAGCGATAGTT 6

```

Search completed: February 16, 2006, 18:23:48
Job time : 435.034 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 16, 2006, 17:47:13 ; Search time 297.143 Seconds
(without alignments)
792.964 Million cell updates/sec

Title: US-10-073-293A-6

Perfect score: 552

Sequence: 1 MSFEVLLGLLVGVNYCFR.....IPTLLSALAYGLAWKMAII 111

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 7204252 seqs, 1061369211 residues

Total number of hits satisfying chosen parameters: 14408504

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-O=/abs/ABSSWEB spool/US10073293/runat 15022006 120743 1175/app_query.fasta.1
-DB=Published Applications NA New -QFMT=fastap -SUFFIX=p2n.rnpbn -MINMATCH=0.1
-LOOPEU=0 -LOOPEXT=0 -UNITS=bases -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTPMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -HOST=abs03p
-USPR=US10073293 @CGN 1.1 579 @runat 15022006 120743 1175 -NCPU=6 -ICPU=3
-NO MMAP -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA New:

- 1: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
- 11: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
- 12: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	78	14.1	1400	12	US-11-136-527-6758
C 2	78	14.1	1879	12	US-11-136-527-2662
C 3	76	13.8	4041	8	US-10-793-626-3573
C 4	76	13.8	4342	8	US-10-793-626-3989

5	74	13.4	729	11	US-11-082-389-363	Sequence 363, App
6	72.5	13.1	641	6	US-09-925-065A-848859	Sequence 848859, App
7	72.5	13.1	3920	8	US-10-821-234-254	Sequence 254, App
C 8	72.5	13.1	100000	12	US-11-124-367A-5084	Sequence 5084, App
9	71.5	13.0	340	6	US-09-925-065A-921304	Sequence 921304, App
10	71.5	13.0	342	6	US-09-925-065A-943123	Sequence 943123, App
C 11	71.5	13.0	343	6	US-09-925-065A-924427	Sequence 924427, App
C 12	71.5	13.0	344	6	US-09-925-065A-922200	Sequence 922200, App
C 13	71.5	13.0	632	6	US-09-925-065A-482686	Sequence 482686, App
14	70	12.7	1632	12	US-11-052-554A-546	Sequence 546, App
15	69.5	12.6	653	6	US-09-925-065A-874878	Sequence 874878, App
C 16	68	12.3	3994	12	US-11-136-527-210	Sequence 210, App
17	67.5	12.2	1673	12	US-11-136-527-2344	Sequence 2344, App
18	67	12.1	3105	8	US-10-392-234A-29	Sequence 29, Appl
19	66.5	12.0	532	6	US-09-925-065A-808746	Sequence 808746, App
20	66.5	12.0	532	6	US-09-925-065A-859760	Sequence 859760, App
21	66.5	12.0	534	6	US-10-793-626-2531	Sequence 2531, App
C 22	66.5	12.0	538	6	US-09-925-065A-778471	Sequence 778471, App
23	66.5	12.0	695	8	US-10-453-372-889	Sequence 889, App
C 24	66.5	12.0	1593	8	US-10-995-805-3	Sequence 3, Appli
C 25	66.5	12.0	2401	6	US-09-925-065A-90971	Sequence 90971, A
C 26	66.5	12.0	3229	8	US-10-793-626-3867	Sequence 3867, App
C 27	66.5	12.0	37500	8	US-10-522-037-1	Sequence 1, Appli
C 28	66	12.0	639	6	US-09-925-065A-385624	Sequence 385624, App
C 29	66	12.0	683	6	US-09-925-065A-678535	Sequence 678535, App
C 30	66	12.0	683	6	US-09-925-065A-678536	Sequence 678536, App
C 31	66	12.0	1269	12	US-11-098-686-9055	Sequence 9055, App
C 32	66	12.0	100000	12	US-11-124-367A-5024	Sequence 5024, App
C 33	66	12.0	394488	8	US-10-995-561-13473	Sequence 13473, A
C 34	66	12.0	1457619	12	US-11-098-686-8739	Sequence 8739, App
C 35	65.5	11.9	2550	12	US-11-136-527-2113	Sequence 2113, App
C 36	65	11.8	201	8	US-10-995-561-77301	Sequence 77301, A
C 37	65	11.8	1868	9	US-11-072-512-1745	Sequence 1745, App
C 38	65	11.8	98309	12	US-11-124-368A-2921	Sequence 2921, App
39	64.5	11.7	556	8	US-10-453-372-895	Sequence 895, App
40	64.5	11.7	593	8	US-10-453-372-893	Sequence 893, App
41	64.5	11.7	635	6	US-09-925-065A-805065	Sequence 805065, App
C 42	64.5	11.7	630	6	US-09-925-065A-747940	Sequence 747940, App
43	64.5	11.7	631	6	US-09-925-065A-805092	Sequence 805092, App
44	64.5	11.7	655	8	US-10-453-372-891	Sequence 891, App
45	64.5	11.7	739	6	US-09-925-065A-879702	Sequence 879702, App

ALIGNMENTS

RESULT 1
US-11-136-527-6758/c
; Sequence 6758, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 6758
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-6758

Alignment Scores:	8.28	Length:	1400
Pred. No.:	78.00	Matches:	34
Score:	46.4%	Conservative:	17
Best Local Similarity:	30.9%	Mismatches:	34
Query Match:	14.1%	Indels:	25
DB:	12	Gaps:	5

1

```

Db 434 ATCATGGTGTCTGTGTTGGCGGATTCATAGCGCAGGACATGGCGTTTATTAAAGCAGCATG 493
Qy 75 ValGlyPheAlaValLeuGlyAlaSerPheTyrLys-----ThrArgSerIleIle 91
Db 494 ---GCCTTTGCTCTGCGGCTGTGCTGTCTTCGTATGCGCTCTTCGTGTTCGCATGATGATT 550
Qy 92 IleProThrLeuLeuSerAlaLeu 99
Db 551 ATCCCTGCACAAATGTTCCCTGCTT 574

RESULT 6
US-09-925-065A-848859
; Sequence 848859, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 848859
; LENGTH: 641
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-848859

Alignment Scores:
Pred. No.: 16.2 Length: 641
Score: 72.50 Matches: 31
Percent Similarity: 40.0% Conservative: 13
Best Local Similarity: 28.2% Mismatches: 37
Query Match: 13.1% Indels: 29
DB: 6 Gaps: 6

US-10-073-293A-6 (1-111) x US-09-925-065A-848859 (1-641)
Qy 4 GluValLeuLeuLeuGlyLeuValGlyValAlaAsnTyrCysPheArgTyrLeuPro 23
Db 213 GAGGTAAATGAAGCTGGACCCCTTGTAGGGGCTGGC-----CCC 251
Qy 24 LeuArgLeu-----ArgValGlyAsn-----AlaArg 32
Db 252 CTGGGGGTCTTCTTGGGGGTAGGACGACAGAGTGAGAGGGGGCAGTGGCCAGA 311
Qy 33 ProThrLysArgGlyAlaValGlyIleLeuLeuAspThrIleGlyIleAlaSerIleCys 52
Db 312 GCACCTTCTGAAGGTAAATCTCGGGTGGCGATGACAGGGTGGGCTGAGGGGGCACTGC 371
Qy 53 AlaLeuLeuValSer-----ThrAlaProGluValMetHisAsp 66
Db 372 TCCCTAGTCTATGATCACCACCGAGTTTACCTCCCACTGCTTTGTCTTGAAGCAGAT 431
Qy 67 ThrArgArgPheValProThrLeuVal-----GlyPheAlaValLeuGlyAlaSerPhe 84
Db 432 TTGAAT-----CCTGCTCTTCCCTGGCAGCCAGACAGCCCTTGGCAAGCATTT 482
Qy 85 TyrLysThrArgSerIleIleIleProThr 94
Db 483 TACCTGAGACCCCTCCACTGTGCTGCTACC 512

```

```

RESULT 7
US-10-821-234-254
; Sequence 254, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 254
; LENGTH: 3920
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-821-234-254

Alignment Scores:
Pred. No.: 151 Length: 3920
Score: 72.50 Matches: 28
Percent Similarity: 52.9% Conservative: 17
Best Local Similarity: 32.9% Mismatches: 31
Query Match: 13.1% Indels: 9
DB: 8 Gaps: 7

US-10-073-293A-6 (1-111) x US-10-821-234-254 (1-3920)
Qy 2 SerTyrGluValLeuLeuLeuGly-----LeuLeuValGlyVal-----AlaAsnTyrCysPhe 19
Db 588 TCGGTTGAAATGCTGATCTCGGTCGCTTGCTTATTGGCCTCTTCTCGGGACTCTGCACA 647
Qy 20 ArgTyrLeuProLeuArgLeuArgValGlyAsnAlaArgProThrLys---ArgGlyAla 38
Db 648 GGTTTGTGGCC-----ATGTACATTGGAGAGATCTCGCCTACTGCGCTCGGGGTGCC 701
Qy 39 ValGlyIleLeuLeuAspThrIleGlyIleAlaSerIleCysAlaLeuLeuValSer 58
Db 702 TTTGGC---ACTCTCAACAGCTGGGCATC-----GTTGTGGAATCTCTGTGGCCAG 752
Qy 59 Thr---AlaProGluValMetHisAspThrArgPheValProThrLeuValGlyPhe 77
Db 753 ATCTTTGGTCTGGAATTATCATCTTGGGTCTGGAAGACTATGCGCGCTACTACTGGGTTT 812
Qy 78 AlaValLeuGlyAla 82
Db 813 ACCATCTTCTCTGCT 827

RESULT 8
US-11-124-367A-5084/c
; Sequence 5084, Application US/11124367A
; Publication No. US20060024700A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: Hongjin Huang
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; Fibrosis Methods of Detection and Uses Thereof
; FILE REFERENCE: CLO01519.ORD
; CURRENT APPLICATION NUMBER: US/11/124,367A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,846
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/582,609
; PRIOR FILING DATE: 2004-06-25
; PRIOR APPLICATION NUMBER: US 60/599,554
; PRIOR FILING DATE: 2004-08-09
; NUMBER OF SEQ ID NOS: 34460
; SOFTWARE: FastSeq for Windows Version 4.0

```

```
; SEQ ID NO 5084
; LENGTH: 100000
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-124-367A-5084

Alignment Scores:
Pred. No.: 8.17e+03 Length: 100000
Score: 72.50 Matches: 31
Percent Similarity: 42.5% Conservative: 14
Best Local Similarity: 29.2% Mismatches: 33
Query Match: 13.1% Indels: 28
DB: 12 Gaps: 2

US-10-073-293A-6 (1-111) x US-11-124-367A-5084 (1-100000)

Qy 1 MetSerTyrGluValLeuLeuGlyLeuValGlyValAlaAsnTyrCysPheArg 20
Db 78013 CTGCTTATGAACTGGGCTGCTTGGTCTCTTAGATTCTCAGCACCTCTTCAATTGAGA 77954
Qy 21 TyrLeuProLeuArgLeuArgValGlyValAsnAlaArgProThrLysArgGlyAla-ValG1 40
Db 77953 GGGTGTGCTGCTCCCTCTTGTGCCACAGCTAGGAACCTCTCAAGACGAGCGAGTTGG 77894
Qy 40 YlleLeuLeuAspThrIleGlylleAlaSerlleCysAlaLeuValValSerThrAl 60
Db 77893 GGCAGCCATGGGTGACTTAGTCTCTCGTCT----- 77863
Qy 60 aProGluValMetHisAspThrArgArgPheValProThrLeuValGlyPheAlaValle 80
Db 77862 -----CCAGGAGTACTGCTCTTCATTGCTCAGTCCAGTGTCTT 77822
Qy 80 uGlyAlaSerPheTyrLysThrArgSerllelleilleProThrLeuLeuSerAlaLeuAl 100
Db 77821 AGGACTC-----TGGGTTTGTCTGTTTGT 77795

Qy 100 aTyrGlyLeuAlaTrp 105
Db 77794 TTTTGGTTTGTGTTGG 77779

RESULT 9
US-09-925-065A-921304
; Sequence 921304, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 921304
; LENGTH: 340
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-921304

Alignment Scores:
Pred. No.: 10 Length: 342
Score: 71.50 Matches: 17
Percent Similarity: 52.9% Conservative: 20
Best Local Similarity: 24.3% Mismatches: 18
Query Match: 13.0% Indels: 15
DB: 6 Gaps: 2

US-10-073-293A-6 (1-111) x US-09-925-065A-921304 (1-342)

Qy 51 IleCysAlaLeu-----LeuVal 56
Db 95 GTCTGTGCCCTCGAGGAGTTGCAAACTTAGTCTCGGGAATAAGACAAATRCAGTGTAAAT 154
Qy 57 ValSerThrAlaProGluValMetHisAspThrArgArgPheValProThrLeu---Val 75
Db 155 ACGACTACAAGGAAGAAAGTGAATAATGGCACAAGCAGGTGTATCGACCTTGACTGCA 214
Qy 76 GlyPheAlaValLeuGlyAlaSerPheTyrLysThrArgSerllelleilleProThrLeu 95
Db 215 GGTGTGAATCACCTGGGAGCTTTTAAAAACACCATGACCAGGTCTCTTCCACAGAGTT 274

Best Local Similarity: 24.3% Mismatches: 18
Query Match: 13.0% Indels: 15
DB: 6 Gaps: 2

US-10-073-293A-6 (1-111) x US-09-925-065A-921304 (1-340)

Qy 51 IleCysAlaLeu-----LeuVal 56
Db 93 GTCTGTGCCCTCGAGGAGTTGCAAACTTAGTCTCGGGAATAAGACAAATRCAGTGTAAAT 152
Qy 57 ValSerThrAlaProGluValMetHisAspThrArgArgPheValProThrLeu---Val 75
Db 153 ACGACTACAAGGAAGAAAGTGAATAATGGCACAAGCAGGTGTATCGACCTTGACTGCA 212
Qy 76 GlyPheAlaValLeuGlyAlaSerPheTyrLysThrArgSerllelleilleProThrLeu 95
Db 213 GGTGTGAATCACCTGGGAGCTTTTAAAAACACCATGACCAGGTCTCTTCCACAGAGTT 272
Qy 96 LeuSerAlaLeuAlaTyrGlyLeuAlaTrp 105
Db 273 CTGATTTTCATTGGTCTGGGTACAGCCTGG 302

RESULT 10
US-09-925-065A-943123
; Sequence 943123, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 943123
; LENGTH: 342
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-943123

Alignment Scores:
Pred. No.: 10 Length: 342
Score: 71.50 Matches: 17
Percent Similarity: 52.9% Conservative: 20
Best Local Similarity: 24.3% Mismatches: 18
Query Match: 13.0% Indels: 15
DB: 6 Gaps: 2

US-10-073-293A-6 (1-111) x US-09-925-065A-943123 (1-342)

Qy 51 IleCysAlaLeu-----LeuVal 56
Db 95 GTCTGTGCCCTCGAGGAGTTGCAAACTTAGTCTCGGGAATAAGACAAATRCAGTGTAAAT 154
Qy 57 ValSerThrAlaProGluValMetHisAspThrArgArgPheValProThrLeu---Val 75
Db 155 ACGACTACAAGGAAGAAAGTGAATAATGGCACAAGCAGGTGTATCGACCTTGACTGCA 214
Qy 76 GlyPheAlaValLeuGlyAlaSerPheTyrLysThrArgSerllelleilleProThrLeu 95
Db 215 GGTGTGAATCACCTGGGAGCTTTTAAAAACACCATGACCAGGTCTCTTCCACAGAGTT 274

Alignment Scores:
Pred. No.: 9.96 Length: 340
Score: 71.50 Matches: 17
Percent Similarity: 52.9% Conservative: 20
```

632
17

Percent Similarity:	52.9%	Conservative:	20
Best Local Similarity:	24.3%	Mismatches:	18
Query Match:	13.0%	Indels:	15
DB:	6	Gaps:	2

US-10-073-293A-6 (1-111) x US-09-925-065A-482686 (1-632)

QY		51	IleCysAlaLeu-----LeuVal 56
			::::
Db		248	GTCCTGGCCCTCGAGGAGTGCACAACTTAGTCTCGGAATAAGACAATGCAGTGTAATT 189
			:::
QY		57	ValSerThrAlaProGluValMetHisaspThrArgArgPheValProThrLeu---Val 75
			:::
Db		188	ACGACTACAGGAGAAGAGTCAATAATGGCCAAAGCAGTGGTTATGCACTTGCATGTCCA 129
			:::
QY		76	GlyPheAlaValLeuGlyAlaSerPheTyriLysThrArgSerIlellelleProThrLeu 95
			:::
Db		128	GTTGGAAATCACCTGGGGAGCTTTAAAACACCAGTACCAGGGTCTCTCTCCACAGAGTT 59
			:::
QY		96	LeuSerAlaLeuAlaTyrGlyLeuAlaTrp 105
			:::
Db		68	CTGATTTCATTGGTCTGGGGGTACAGCTGG 39
			:::

RESULT 14

US-11-052-554A-546
; Sequence 546, Application US/11052554A
; Publication No. US20050288866A1

```

: GENERAL INFORMATION:
: APPLICANT: Sachdeva, et al.
: TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
: TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
: FILE REFERENCE: 30853/40359A
: CURRENT APPLICATION NUMBER: US/11/052,554A
: CURRENT FILING DATE: 2005-02-07
: PRIOR APPLICATION NUMBER: US 60/589,227
: PRIOR FILING DATE: 2004-07-20
: PRIOR APPLICATION NUMBER: IN 173/DEL/2004
: PRIOR FILING DATE: 2004-02-06
: NUMBER OF SEQ ID NOS: 763
: SOFTWARE: PatentIn version 3.3
: SEQ ID NO 546
: LENGTH: 1632
: TYPE: DNA
: ORGANISM: Mycobacterium tuberculosis H37Rv
US-11-052-554A-546

```

Alignment Scores:		
Pred. No.:	108	Length: 1632
Score:	70.00	Matches: 24
Percent Similarity:	38.0%	Conservative: 5
Best Local Similarity:	32.0%	Mismatches: 26
Query Match:	12.7%	Indels: 20
DB:	12	Gaps: 2

US-10-073-293A-6 (1-111) x US-11-052-554A-546 (1-1632)

38	AlaValGlyIleLeuLeuAspThrIleGlyIleAlaSerIleCysAlaLeuLeuVal---	56
692	CGCGTTTCGCTGCTTTTGGAAACCGCGGCGCGCAACGCGCGCGCAGCGGTTCCG	751
	:::	
57	-----ValSerThrAlaProGluValMetHisaspThrArgArgPheValProThrLeu	74
		:::
752	TCTTTGGCGTCGCGGCGCGCGCGCTCGCGCGATCGAGCGAAGACGTCGCGCGGCACGC	811
75	ValGlyPheAlaValLeuGlyAlaSerPheTyrIlyThrArgSerIleIleIleProThr	94
812	CGCGGCGCGCGCGCTCGCGCGC-----	835
95	LeuLeuSerAlaLeuAlaTyrGlyLeuAlaTriPlysValMetAla	109
836	-----GTCTATTCCTGGCTCGCGCGGTATGCGC	865

RESULT 15

US-09-925-065A-874878
; Sequence 874878, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:

```

; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT FILING DATE: 2001-08-08
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 874878
; LENGTH: 653
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-925-065A-874878

Alignment Scores:
Pred. No.: 40.3 Length: 653
Score: 69.50 Matches: 16
Percent Similarity: 61.0% Conservative: 9
Best Local Similarity: 39.0% Mismatches: 15
Query Match: 12.6% Indels: 1
DB: Gaps: 1

US-10-073-293A-6 (1-111) x US-09-925-065A-874878 (1-653)

```

Alignment Scores:		
Pred. No.:	40.3	Length: 653
Score:	69.50	Matches: 16
Percent Similarity:	61.0%	Conservative: 9
Best Local Similarity:	39.0%	Mismatches: 15
Query Match:	12.6%	Indels: 1
DB:	6	Gaps: 1

US-10-073-293A-6 (1-111) x US-09-925-065A-874878 (1-653)

	QY	10	IeuIeuValGlyValAlaAsnTyrCysPheArgTyrIeuProIeuArgIeuArgValGly	29
			::: ::: ::: ::: :::	
	Db	200	TGTITTAATGGGATCACCCAGC--TGTTTTACTCTCCCATCTCTTCGAGGTGACT	256
	QY	30	AsnAlaArgProThrLysArgIeValGlyIleLeuLeuAspThrIleGlyIleAla	49
			::: ::: ::: ::: ::: :::	
	Db	257	ACAGAAGAGGACAAAAAACGGATGCAGTGGGACTCACTTTTAGATCATCTGGGACTACTG	316
	QY	50	Ser	50
	Db	317	AGC	319

Search completed: February 16, 2006, 18:26:50
Job time : 309.143 secs

This Page Blank (uspto)